

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 05:31:50 ; Search time 2313.83 Seconds
(Without alignments)
16279.401 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800
Sequence: 1 atgctctcttcacgcgcaac.....tggcgaagcaggtcgtgtga 1800

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	AX148306	1800	100.0	1800	100.0	3869	AX148306
2	AX148306	1800	100.0	52232	8	AB026549	AB026549 Arabidops
3	AX148306	1800	100.0	2331	8	AB026517	AB026517 Arabidops
4	AX148306	1800	100.0	2331	8	AY056255	AY056255 Arabidops
5	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
6	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
7	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
8	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
9	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
10	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
11	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
12	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
13	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
14	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
15	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
16	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
17	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
18	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
19	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
20	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
21	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
22	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
23	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
24	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
25	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
26	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
27	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
28	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
29	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
30	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
31	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
32	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops
33	AX148306	1800	100.0	2331	8	AC013430	AC013430 Arabidops

ALIGNMENTS

RESULT	LOCUS	SEQUENCE	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	CDs
1	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
1	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
2	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
3	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
4	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
5	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
6	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
7	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
8	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
9	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
10	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
11	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
12	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
13	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
14	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
15	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
16	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
17	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
18	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
19	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
20	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
21	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
22	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
23	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
24	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
25	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
26	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
27	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
28	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
29	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
30	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
31	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
32	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	
33	AX148306	Sequence 5 from Patent EP116794.	AX148306.1	GI:14347193	thale cress.	1 (bases 1 to 1800)	Iuchi, S., Kobayashi, M. and Shinozaki, K.	Transgenic plants carrying neoxanthin cleavage enzyme gene	Riken (JF)	Location/Qualifiers	

Location/Qualifiers
1. 1800
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
1. 1800
/note="unnamed protein product"

```
/codon_start=1
/protein_id="CAC41194.1"
/db_xref="GI:14347194"
/translation="MASPTAAVSGRMVIGGNHOPPLSYSSOSSDLVSYSSSLPMASRV
TRKINVSALHTPPALHPFPKROSSNSPALVYVPRKESITKONMLFORAAAALDAABG
FLVSHKEHLPLKPADPSVQIAGNEAPVNEPVRKNLPLVSKLPDLSIGVYRNGANP
LHEPVTGHHFDFDGMVAHVEHGSASVACRTQTNFVEVEROLGRVFPKALIGELH
GHTGILARMLFARAAGVDPAPAGTVANAGLVFNGLRILAMSDDLPOVOITPNG
DKTVGRDFDGFESTMIAHPRVDESEGLFALSYDVSKPYLYKFEFSPDGRKSPD
VEIOLDPTMMHDFALITENFVYVDPDOVFELPRMIRGSGSPVYVYDKNVARPGLIDKY
AEDSSNIMKIDAPRCFCFHLMMANDEPTEDEVYVIGSMTPPDSIFNSDENLSVLS
EIRLNLTGSESTRNPDIISNEDQVNLLEGVMNRMNLGKRTFAVYLAELPWPXVSGFA
KIVLTGVEKHSHLYGDMRYGGEPLFLPEGEDEDEGYTILCEVHHEKTKWSLQIVNAV
SLEVEATVKLSRPVPGFHFHTGFIADDLAKOV"
```

BASE COUNT 458 a 464 c 439 g 439 t

ORIGIN

Query Match 100.0%; Score 1800; DB 6; Length 1800;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggctctttcaaggcaagcgctcggttcttggaagatggtcgtggtgcaataact 60
DB 1 ATGGCTTCTTTCACGGCAACGGCTCGGTTCTGGAGATGGCTTGGCAATCATACT 60

OY 61 cagcgcacatctcgtctctcacaagctccgacttgagatctatgactcctactatg 120
DB 61 CAGCGGCCATTATCGCTTTCTAAAGCTCCGACTTGAGTTATGACTCTTACCTTANG 120

OY 121 gccagtcgtgtacacgyltaagctcaatgyltcatctgcgtcctcaactcctcagctt 180
DB 121 GCCAGTCGTCACACGTAGCTCAATGTTTCATCGCGCTTCACACACTCTCCAGCTTT 180

OY 181 cattccctcaagcaatcacaactcctccgcacatgtgtgtaagcccaaaagcaaga 240
DB 181 CATTCCCTCAAGCAATCACTAACTCTCCGCAATGTTTAAAGCCCAAAAGCAAGAA 240

OY 241 tccaacactaaacagatgattgttccagagaagcgcgcgagcagcgtttgacgcgag 300
DB 241 TCCAACTAAACAGATGATTTGTTCCAGNAGCGCGCGGCGACGCTTGGACCGCGGG 300

OY 301 gagggttccctgtcagccacgagaagctacacccgcttccctaaacggcgtatccctagt 360
DB 301 GAGGGTTTCTGTGACGCCACGAGAAAGCTACACCGCTTCTTAAAAAGGCTATCTAGT 360

OY 361 gttcagatccgcgcgaatatttgtcgcgtggaatgaacagccgcctgcggtatattccg 420
DB 361 GTTCAGATCCGCCGAATTTTGTCTCGGTGAATGAACACCCGCTCGGCGTATCTTCCG 420

OY 421 gtgtcgcgaaaaactcccgatccatccaaagagtgatgtgcgcgaacgcgaactaaccca 480
DB 421 GTGTGCGGAAAACTTCCCGATTCATCAAGAAGAGTGTATGTCCGCAACGCACTTAACCCA 480

OY 481 cttaacgagcgggtgaacaggtacacactcttcgacgagagacggtatggttcaagcgcgtc 540
DB 481 CTTCACGAGCCCGGTGACAGGTCACCACTTCTTCGACGAGAGAGGTATGTTTACCCCGTC 540

OY 541 aaatccgaacaggttcagctagctacgcttgcgcgtttactcaactcaaacgggtttgt 600
DB 541 AAATTCGAACAGGTTTCACTAGCTACGCTTGC CGGTTTACTCAGACTAACCGGTTTGT 600

OY 601 caggaacgtlcaatltggttcgacgcggttttccccaagccatcggtgagcttcaagcgcac 660
DB 601 CAGGAACGTCAATGTGGTGCACCGGTTTCCCCAAAGCATGGGTGAGTTTACCGGCCAC 660

OY 661 accggtattgcgcgaactactatcttaagcagagctgcaagcggatagtcgacccg 720
DB 661 ACCGATATTGCGCGACTACTATTTCTACGCCAAGCTGACGCGGATATGTCGACCCG 720

OY 721 gcaacgcgaacccggtgttagctaaacgcggttggctatattaaagcgcggtattggt 780
DB 721 GCACACGGAACCGGTGTAGCTAAACCGCGTGTGCTATTTCATAGCCGGTATTGCT 780

```
OY 781 atctcggaagatgatttaactcttaccgaagtctcagatcaactcccatggagatttaaac 840  
DB 781 ATCTCGGAAGATGATTTAACCTTACCAGATTCACAGTTCACACTCCCAATGGAGATTAAAAAC 840

OY 841 gtgtgtcgttgcgatttgaatgaacatagatccacaatgattgtgccaccggaagtc 900  
DB 841 GTGTGTCGTTGCGATTTGATGAGACAATTAGAATCCACAAATGATTGGCCACCCGAAGTTC 900



OY 901 gaccggaaatccggtggaactcttcgcttaagctacagagctcggttccaagcctta 960  
DB 901 GACCGGAAATCCGGTGAACCTTCTGTTTAACTTCAGAGAGCTGTCGTTTCAAGGCTTAACCTA 960



OY 961 aaatacttcgattcttcacgcgaggaactaaataacccgagctcgagatctgaacttgat 1020  
DB 961 AAATACTTCGATTTCTCACCGGACGGAACCTTAATACCGGAGCTGAGATTAGCTTGAT 1020



OY 1021 cagccaacgattgatacagatttcgcatatcagagaacattcgtcgtctacactgacag 1080  
DB 1021 CAGCCAACGATGATGACAGATTTCCGGAATTCAGAGAACTTGTGTCGTACTGACCCAG 1080



OY 1081 caagtcggttttcaagctgcgcggaagatgatacgcggttgggtctccggttgaacag 1140  
DB 1081 CAAGTCGTTTAACTGCGCGGAGATGATCCGCGGTGCTCCGCTGCTTACGACAAG 1140



OY 1141 aacaaggtcgaagatttcggaattttagacaataacgcgaagattcactcgaacattag 1200  
DB 1141 AACAAAGTTCGAAGATTCGGGATTTTAAACAATACCGGAAGATTTCATCGAATTAAG 1200



OY 1201 tggattgatactcagatgtcttcgtcctcctcctggaacgcttgggaagagccaga 1260  
DB 1201 TGGATTGATGCTCCAGATGTTGCTTGTCTTCCATCTGTGGAAGCTTGGGAAGACAGCA 1260



OY 1261 acaagatgaagtcgtctgataaggtctctgtatagatccacccagactcaattcaagag 1320  
DB 1261 ACAGATGAAGTCTGTGTAGGTCCTGTATGACTCCACCGAGCTCAATTTTCAACGAG 1320



OY 1321 tctgaacgaatctcaagagtgctcctgtctgaaatccgccttaactcacaacgcgttga 1380  
DB 1321 TCTGACGAATCTCAAGAGTGTCTGTGTGAATCCGCTCAATCTCAATCCAGGTGA 1380



OY 1381 tcaactcgcgttcgatactctccaaacgaagatcaacaagctcgaacgagagtg 1440  
DB 1381 TCAACTCGCGCTCGCATCTCTCCAACGAAGATCAACAAGTCAACCTCGAACACAGGATG 1440



OY 1441 gftcaacagaacatgctgcgcgttaaaccaaatccgcttacttggctttagcgaagcgt 1500  
DB 1441 GTCAACAGAAATGCTCGCGGCTAATAACCAAAATTCGCTTACTTGCTTTAGCCGACCG 1500



OY 1501 tggcctaaagctcagagattcgtctaagtgtatctcaactactcgtgagaagttaagaacat 1560  
DB 1501 TGGCTTAAAGTCTCAGAGATTCGCTTAAAGTTGATCTCAGTACGAGAGATTAAAGAACT 1560



OY 1561 ctttaagcgaataacccgttaacgagagagagcctcgttcttcccggaagagagagag 1620  
DB 1561 CTTTACGGGATTAACCGTTTACGAGAGAGCCCTCTGTTTCTCCCGGAAGAGAGAGAG 1620



OY 1621 gaaagcgaagatacatcctcgttctcgttcaacgagcgaagaagacatgaaatcgaagta 1680  
DB 1621 GAAAGCAAGAGATCATCTCTGTTTCTGTTTCAACGAGAGAAACATGAAATCGGAGTTA 1680



OY 1681 cagatagttaaacgcgtttagcttgaagttgaagaaacggtttaaacttcgltcaaggtt 1740  
DB 1681 CAGATAGTTAAACGCGTTAGCTTAGAGGTTGAAGCAACGGTTAAACTTCCTCCGCAAGGGTT 1740



OY 1741 ccgtacggaattcaacggtatcaatcctcggagccgagatgttggcgaagcaggtcgtgtga 1800  
DB 1741 CCGTACGGATTTCAACGATTAATTCATTCGAGCCGATGATTTGGCCAAACGAGCTCGTGTGA 1800



RESULT 2  
AB026549 3869 bp DNA linear PLN 24-OCT-2001  
LOCUS


```

DEFINITION Arabidopsis thaliana gene for neoxanthin cleavage enzyme, complete cds.
ACCESSION AB026549
VERSION AB026549.1 GI:16416373
KEYWORDS neoxanthin cleavage enzyme.
SOURCE Arabidopsis thaliana (strain:Col.) DNA.
ORGANISM Arabidopsis thaliana
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots: Rosidae: eustosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 3869)
AUTHORS Iuchi,S., Kobayashi,M. and Shinozaki,K.
TITLE Characterization of neoxanthin cleavage enzyme from Arabidopsis thaliana
JOURNAL Unpublished
AUTHORS Iuchi,S. and Shinozaki,K.
TITLE Direct Submission
JOURNAL Submitted (22-APR-1999) Satoshi Iuchi, RIKEN, Plant Mol. Bio.; 3-1-1 Koyadai, Tsukuba 305-0074, Japan
FEATURES
source location: Iuchi et al. riken-go.jp, Tel: 81-298-36-4359)
1..3869
/organism="Arabidopsis thaliana"
/strain="Col."
/db_xref="taxon:3702"
1831..3630
/codon_start=1
/product="neoxanthin cleavage enzyme"
/protein_id="BAB70609.1"
/db_xref="GI:16416374"
/translation="MASFTATAAVSGRWLGSGNHPDLPSSOSSDLSYCSLPMASRV
TRKLNSSALHTPPALHPKROSSNSPALIVKPKRESNTOMNLFQAAAAALDAAG
FLYSHEKHLPTKADPSVOIAGNFAPVNPQPRNRPVYGKLPDSIKYIVRYRANG
LHPVYGHFFDGDGVHAYKFEHGSASVACRFQINRFYQERDLPKVPKALGELH
GHTGIARLMFLYAAAGIVDPAPGTVANAAGLVFNGRLAMSEDLPOVQTTTPG
DLKTVGRFDFDQLESTIMIAHPKVPDSEGLFALSYVNVSKPYLKFYRFSDDGTSYD
VEIOLDPRTMMHDFAITENFVVDQOVFKLPMIGSGSPVYDKKVAEFGILNDY
AEDSSNKKWIDAPCFEHLNMAEPEPTGVVYIGSCMTPDPSIFNESPENKSVLS
EIRLNKGTSESTRPITSNEDOVNLEAGVNNRMRLRKRFKALALAEPRPKSGRA
KVDLTITSEVKHHLXGDKRNYGEBFLPEBGEDEBEGITLCEHDEKTKWSLQIVNAV
SLEEVATVKLPSPRVYGFHGTFIGADDLAQV"

BASE COUNT 1155 a 845 c 773 g 1096 t
ORIGIN

Query Match 100.0%; Score 1800; DB 8; Length 3869;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggctcttcacggcaagcgctgcggttctcgtggaagatggtctgtggaatcact 60
|||||
DB 1831 ATGGCTTCTTCACGGCAAGCGCTGCGGTTCTGGAGATGCGTTGGCAATCATCT 1890
QY 61 cagcgccatcatcgctcttcctcaagctcgaactgaatattgttagctccttaact 120
|||||
DB 1891 CAGCGGCATTAATGCTTTCACAAAGCTCGCACTTGAATATGAGCTCTTACTATG 1950
QY 121 gccagctgtgcacagcaagctcaatgttctcattcgtcgtctcaacccctccagctct 180
|||||
DB 1951 GCCAGTCTGTGCACAGCTCAATGTTTCACTGCGCTTCACACCTCCACACTCTT 2010
QY 181 cattccctaagcaatcatcaaacctccgcgcatgtgtgttaagcccaagaagcaaa 240
|||||
DB 2011 CATTTCCCTAAGCAATCATCAAACTCTCCGCCATGTGTGTTAAGCCCAAGCCAAATA 2070
QY 241 tccaacactaaacagatgaattgttccagaagcgcgcgcgcaagcgttgaagcgagcg 300
|||||
DB 2071 TCCAACACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGGCGAGCTTGAGACGCGCG 2130
QY 301 gaaggttccctgtcagcaagcaagctacacccgcttcttaaaagcgctgactcagt 360
|||||
DB 2131 GAGGTTTCTTGTGCAGCAGAGAAAGCTACACCGCTTCTTAAACGCGTGATCTCTAGT 2190

QY 361 gtccagatcccggaatttcgtcccggtgaatgaacacccggtccggtcaatcttcg 420
|||||
DB 2191 GTTCAGATCGCCGGAATAATTTTGTCTCGGTGAATGACAGCCGCTCGGCGTATTTCCG 2250
QY 421 gtgtcggaaaactcccgatccatcaaaagagtgtatgtgtcgcaacggaactaacc 480
|||||
DB 2251 GTGGTCGGAACACTTCCGATTTCCATCAAGAGAGTGTGTGGCAGACGAGACTTAACCA 2310
QY 481 cttaacgagccggtgaacaggttcacacactctctcgacggagaagcgtatgttcaagcgtc 540
|||||
DB 2311 CTTCAAGACCGGTGACAGGTCACACCTTCTTCAGACGAGAGCGGTATGTTCACCCGCTC 2370
QY 541 aaattcgacaacggttcagctagctacagcttcggttactcaaacacacggttgtt 600
|||||
DB 2371 AAATTCGAACACGTTCAAGTCACTACCTTCCGCTTACTCAGACTTAACCGGTTGT 2430
QY 601 caggaacgtcaattgtgtcgacacggttctcccaaaagcactcggtgaacttcaacgcccac 660
|||||
DB 2431 CAGGAACGTCAATGGGTGCGACCGGTTTCCCAAGGCATCGGTGACCTTCACGGCCAC 2490
QY 661 accggtattcccgactcatgtctatctcagccagagctgcagccgtatagtgcaccg 720
|||||
DB 2491 AOCGTATTGCCCGACCTCATGCTATTCTACGCCAGAGCTGCAGCCGCTATAGTCACCCG 2550
QY 721 gcaacggaacccggtgtagctaaacgcggttgtgtctatcttaatggccggttatgct 780
|||||
DB 2551 GCACACGGAACCGGTGTAGCTACACCGCGTGTGGCTATTTCAATGAGCGGTTATGGCT 2610
QY 781 atgtcgagagatgaattacatcccaagaatcagaatcaccatcccaatgagatlaaaaaac 840
|||||
DB 2611 ATGTGCGAGATGATTTACCTTACCAACTTCAAGTCACTCCCAATGAGATTTAAAAAC 2670
QY 841 gttggtcggttcgattttgatgagacaattagaatcacaatgattgcccacccgaagctc 900
|||||
DB 2671 GTTGTCTCGCTTCGATTTGATGAGACAAATTAATCAAAATGATGATGCCACCGAAATC 2730
QY 901 gaccgggaatccggtgaactcttcgtcttaagctacagcgctcgtttcaaaagccttaacta 960
|||||
DB 2731 GACCGGAATCCGCTGAACCTCTTCGCTTTAAAGCTACGACGTCGTTTCAAAAGCTTACCTTA 2790
QY 961 aaatacttcgaattctcaccggaacggaataatcaccggaacggtcgaattcagctgat 1020
|||||
DB 2791 AAATCTTCCTCCATCTCTACCGGAGCACTTAATCACCAGCGAGTGCAGATTACAGTTGAT 2850
QY 1021 cagccaacgatagtgcaagatttcggtatcagaagaacttcggtcgttacctgagcag 1080
|||||
DB 2851 CAGCCAAAGATGATGACAGATTTCCGATTCACAGAACTTCGTCGTACCTGACACAG 2910
QY 1081 caagtcgtttcaagctgcggaagatgaatcgcggtgtggtctcgggtgttcaagcaag 1140
|||||
DB 2911 CAAGTCGTTTCAAGCTTCGCGAATGATCCCGGTTGGGCTCCGCGTGTGTACGACAAAG 2970
QY 1141 aacaagtcgaagaattcgggaattttagacaataatcgcggaagattcatcgaacaattaag 1200
|||||
DB 2971 AACAAAGTTCGCAAGATTCGGATTTTACAAATTCGCCGAAGATTTCATCAACATTAAG 3030
QY 1201 tggattgattgtccagatttgccttcgtcttcacatctcttgaaacgcttgggaagagcaaa 1260
|||||
DB 3031 TGGATTGATGCTCCAGATTTGCTTCTGCTTCATCTCTGGAACGCTTGGGAAGAGCAAA 3090
QY 1261 acaagatgaagtcgtctgtaabggttcctgtatgacatccacaagactcaatttcaagag 1320
|||||
DB 3091 ACAAGTGAAGTCGTGTATAGGTCCTGTATGATGATCCACAGACTCAATTTTCAACAGAG 3150
QY 1321 tctgaagaaatctcaagagtgctcgtctcgaatccgcttgaatctcaaaaacggtgaa 1380
|||||
DB 3151 TCTGACGAGATTCACAAAGATGCTCTGTGAATAATCCGCTGAATTCATAAACCGGTGAA 3210
QY 1381 tcaactgcggttcgatcatctccaacgaagaatacaaacggtcaactctgaagcagagatg 1440
|||||
DB 3211 TCAACTCGCGCTCGATATCTTCCAAACGAAGATCAACAACTTCGTAAGCAGAGGATG 3270

[illegible]

TANPVLADADORAHEVREKHWLTKIDAEFOAEDILDELOTEALRRRYVAEAGLGG
LPONIMAGRAIOKRIEPMKMEYVRLLEHHVKHIEVIGLKEYSETREPOWQARSP
DDLPOGRLVGRVEDKIALVNLILSDDEISTIGKPAVISVGMGPQKTTIEIVNDR
VTHEPEYKMMISAGINNVPTVIRAVIADITSSAVNTEDPISLQIOIKLSKREPL
VLDFWESDESSEWFOVATDAEBSIVLTRSEIVTAAKEKTIOMKMLNEEC
WELISRAFGNISVGSINOELGIGKIAQKGLPLAARASHLSKRPMDWYV
SKFSSYTNLSILPVLYKLSYDLPOLKRCFALCSIFPKGHVDFREELVLMALDILY
QPSRSRLNEDIGNDYLDVAOSFPORLDITMTFSFMDLMDLNDLAKAVSGFCRLD
DNIPETSTRHFSFSRSCDASAFRISGICAEPLRTILPNSPTSLISLOLMEKVN
PLNALSGLRLISLHQITNUPKSLGKILRLDLSRRIKKILPEPVCTICLOTL
LISNRDLISLPKSIABINLRLDLVGTPEVMPPEIKLRSQKLSNVIGLSGA
GLHLEKLSHLRGLRLISELQNAFASEAADAGKPKPLFDGLILKKTAVGSPSPS
FNAIACQKQEVRLMEPEHPHILKFCIESIOGAGSPKCSFEGITSVILSSCNLCI
SLPEVQDPLSKYLSIKFNILQVGLDFEGENNSRGVPEOSIOLIKFYGMFMDM
ICPELEDIPLOKILIORCPILRKKPEPLGSTETVTISDCLRAVSGENSFERS
LTPNRESAPRISMSRRELSPPTGNPKSDASTSKOPFASSOSNDNEVYTSLSLS
LPKROJEDDQIETQLGSLPQOEPEEPAVSARSISDIPSLYSMTSTLSVPP
KNEGSIIPGSSSYOYHQYIKSVSPRSSEAIKPSQYDDEDMEXLYKMTLSVPE
LPONLSIHDSCGLSLPENLSEYPNLHELILICHSELESFSGHPPTTLTLYI
RDCKLNFTSLOPTRYSOLEYLFIGSSCNLNPPLFPLKRSLSIRCESEKTF
SIHAGIDDRILALESEIRDCPNLEFPPOGLPPTKLSMILNSCKLOALPEKFLG
TSLISFTIKCPETIETIPGGGFPENLTLCSLCKDKTPRLEMKRLDLENRLNEIDG
GNEDIESFPEGLPKSVFSURISRENLKTLNKGHDTALETMELISGCDKLQISI
DEDLPLISCLIRISSCLTEFAVEVEFEKVLNIPVEIDELIFS"
38132. .41296

CDS

/note="gene_id:MOA2.7"
/codon_start=1
/evidence=not_experimental
/product="disease resistance complex protein"
/protein_id="BAB01339.1"
/db_xref="GI:11994217"
/translation="MTGIGMFLAALFOALFOVLVSEPFRRSEFKREINLERLST
ALLITAVILDAEKOITNPVEKWKWELNDVYHADADILDELRLNIGNESS
SNRLROGRMSIGDFDENSEHELETRLEKVIIRLELAASQNLITKELTAMPKOR
LPPTSLVDESEVGRDDKDEIMAFILPEMKDGIIVAVLIGVIGKATLSOLTN
DOHRSYFGTKVAHVASEEDVEKITRKVESVIRSCFEDLDVLAIVKILKRLTG
LPFLVLDLNMENFADMLRQPIHAAQSQILVITRSRVSIMCAVAMHLOPL
SDGCMGLFMKTYGNOEPLNREIGDAEIVKCGGLPLAVTLGVLVEEKGVE
MERVLSRIMDLPADKSNLLPVLRVSYYPALHRCFVASTIPKGAPEKDVLL
MMAEGFIQORSSKLELIGNEVSELESLSLOKTRHYMHFIMILIOFAGFS
SKPEDECKLOVSEKTRILSLRQNYAEMPEALREKFLTLTPELSTJNSRCCID
QNVSEKLPULRLVLSLHYKILARLPDEFKNIISARFLDSITELKPKSLCTM
YNLOTLILSYCSLKEPLDILNINRYLDLIGTKLRMPRRGRKLSLOTLTFEY
SASGSRISSELGSLHDLHGKLIKIVELRVVDAVDAANLNSKHLEIDVMTGVS
SEENITNPHRTONEAEVEKILPRHILEKIAIEYKGRPMDLSDSFSITVIR
EYOCSTLPSIGLOPLCKELHISGMGLISGRFVSDOOLRODOPRELSLIR
DNLPDMEMLDVRYTRGDLRPLSKLILRPELTGTLPTLPBLISLHITKGLDB
QPDHETSYRLQTLTSSCDTLVKFPLNFAMDLEVDQCTSLVLESNHLRGS
PMLRLNRLINDCOMLPLKLNALPOMLOYITNCRYLROPMEQPOYHHPQHLPRS
NVSGSPKSHGSHRSYSSRYD"
42345. .42872
/note="gene_id:MOA2.8
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01340.1"
/db_xref="GI:11994218"
/translation="MSDLVITAVAFVAVLVFVITLSCIMGSEKDVHSPPLRPGOLR
MPVYRTKDXDGGGTALISVAGTDITYGSGGSGHGGRGRTITSDTKLPSPPPSLP
PRVQRPKSVSRKIDRELIDGTAMLAASSLITGSSASSCSGGGSHHGCGGGGGG
GGGGGGGGGGGGCGG"
complement(join(42980. .43465,43561. .43668,43952. .44119,
44215. .44460,44539. .44677,44759. .44977,45060. .45754))
/note="gene_id:MOA2.9"
/codon_start=1
/evidence=not_experimental
/product="terpene synthase-like protein"
/protein_id="BAB01341.1"
/db_xref="GI:11994219"
/translation="MKPKVRMLKSSKNDSMERIRLILHMLNLGIAVHEFEIDELIC
QARGNDDITAKENDLETISTMEVFRIRGTYMCCSGISGLDSHICHTIYNLADFRT
VDAPNRKGEDGRKESLAEIDRGMLDYEAAHGLGTSESDMALSTTRRLSLVS
NHTATASPHLSKHIONALYRARIENLELVARREYISFEQEDDEHDETLKPAKLNFNY

CDS

CDS

Query Match		100.0%	Score 1800;	DB 8;	Length 52232;
Best Local Similarity		100.0%	Pred. No. 0;		
Matches 1800;		Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Oy	1	atgagcttttcaaggcaacgctgagcttcttggagatgagcttggatgcaatcat	60		
Db	13669	ATGCGTCTTTCACGCGCAACGGCTCGGTTTGGGAATGGCTTGGGCAATATACT	13610		
Oy	61	cagccgcattatcgtctcttcaaaagctcgcagcttgaattatgagctcttacctatg	120		
Db	13609	CAGCGGCATATATCGTCTTTCAAAGCTCCGACTGAGTATTTGATGCTTACCTATMG	13550		
Oy	121	gccagtgatgcaacgcaagctcaatgattatctgcttcaacatctctcagctctt	180		
Db	13549	GCCAGTCTGTGACACGACGACGATGATGTTGATGCGCTTCAACATCTCCACCTGT	13490		
Oy	181	catctccctaagaatcatcaactctccgcgcatgtgtgtaagcccaagaacaa	240		
Db	13489	CATTTCCTTAACATCATCAACATCTCCGCAATGTTGTTAAAGCCCAAGACCAAGAA	13430		
Oy	241	tccaaactaaacagatgaattgttccagagagcgcgcgagcagcttggagcgcg	300		
Db	13429	TCCAACTAAACAGATGAATTTGTTCCAGAGAGGCGGCGGAGCGTTGAGCGGGG	13370		
Oy	301	gagggtttccttgtaagcaagaaagctacacccgcttctcctaaagcgctgaatcctg	360		
Db	13369	GAGGGTTTCCTTGTGACGACGACGACGACGACGCGCTTCAAAACGCTGATCTAGT	13310		
Oy	361	gttcagatcgccggaatattgtctcgggtgaatgaacacgcccgtccgagatcttcg	420		
Db	13309	GTTCAAGATCGCGGAAATTTTCTCGGATGAATGAACACGCGTCCGCGATATCTTCG	13250		
Oy	421	gtgtcggaaaaactcccgatctcaataaagagtgatgtgtgcgaacgagataccca	480		
Db	13249	GTTGCGGAAAACTTCCGATTCATCAAAAGAGGTATGTTGCGCAACGAGACTTAACCA	13190		
Oy	481	cttcacgagcggtgtacaggttaccacttcttcgacgagaaagatgtgttaacgcgtc	540		
Db	13189	CTTCAAGACCGGTTGACAGGTACCACTTCTTGAAGGAGACGATGTTGACCGCGTC	13130		
Oy	541	aaatcgaaacggttcagctagctgacgcttcgagtttaactcaagaactaacggttgtt	600		
Db	13129	AAATTCGAAGACGTTCACTACGTACGCTTCGCGTTTACTCAACATTAACCGGTTTGT	13070		
Oy	601	caggaacgtlcaattggtgtgacggttttcccaaaagcactcgtgagcttcaagcgac	660		
Db	13069	CAGGAACGTCAATTTGGTGACCGGTTTCCCAAGCATGTTGAGCTTACCGGCAC	13010		
Oy	661	accggtattgcccagctcatgtatcttaagccaaagctgtcagccggtatagtcagccg	720		
Db	13009	ACCGGATTATGCCGACTATGTTTACGCAAGCATGACGAGCGCGGTATATCAACCG	12950		
Oy	721	gacacggaacgggtgtagtaacgcggttggctatcttaacgagcggtatctggt	780		
Db	12949	GCACACGGAACGGGTGTACCTACACCGGTTTGGTCTATTTAAGCGCGTATTGGGT	12890		
Oy	781	atgtcggagatgattacacttcaacaaagtlcagatcactcccaatgagattaaaac	840		
Db	12889	ATGTGGAAGATGATTTACCTTACCAAGTTCAGATCCACCAATGAGGATTTAAAAAC	12830		
Oy	841	gttgatggttcgattttgtagaacaattagaatccaatgattgcccacccggaagtc	900		
Db	12829	GTTGATGCTTCGATTGATGATGACAAATTTAGATCCACATGATGTTGCCACCGAAAGTC	12770		

QY 901 gaccggaaatccggtgaactcttcgctttaagtaagaagtcgttccaagcttacccta 960
 |||||
 Db 12769 GACCCGGAAATCCGTTAACTCTTCGTTTAAAGTACAGCTCTCTTCAAGCCCTTACCTA 12710
 QY 961 aatacttcgattctcacgcgcgcgaactaaatcacccgaactcgagattcagcttgat 1020
 |||||
 Db 12709 AATATCTTCGATTCTCACCAGGAGAACTAAATCACCGAGCTCCAGATTCAAGCTTGGAT 12650
 QY 1021 ccgccaagatgatgcagcgaatttcgcgattacagagaacttcgctcgtcgtacactgaacag 1080
 |||||
 Db 12649 CAGCCAAACGATGATCAGCATTTCCGATTTACAGAGAACTCTGCTCGTACCTGACACAG 12590
 QY 1081 caagtgcgttttaagatgcgcgcgcgaatgatcccggtggtgtctcgggtgttttaagcaag 1140
 |||||
 Db 12589 CAAGTCTGTTTCAAGCTGCGCGGAGATGATCCGGGTGGCTCTCGGTGTTTACGCAAG 12530
 QY 1141 aacaagtcgcgaagatccggaattttagacaatatccgcaagatccatcgaacataag 1200
 |||||
 Db 12529 AACAAAGTGCAGAGATTCCGGGATTTTACAAATACGCCGAGATTCAATCGAACATTAAAG 12470
 QY 1201 tggattgattcgcagatttcgtctcgtctccatctctcgtgaacgcttgaggagagacagaa 1260
 |||||
 Db 12469 TGGATTGATGCTCCAGATTGCTCTCTCTCCATCTCTGGAAGCTTGGGAAGAGCAGAA 12410
 QY 1261 acagatgaagtcgtcgtgataggggtccgttatgactccacagactcaatttcaacgag 1320
 |||||
 Db 12409 ACAGATGAAGTCTCGTGTATGGGCTCTGTATGATCTCCACCACTCAATTTTCAACGAG 12350
 QY 1321 tctgaagaagaatcacaagatgctcgtctgtaaatccgcctgaatctcaaaacggttga 1380
 |||||
 Db 12349 TCTGACGAGATTTCAAGAGTGTCTGTCTGAAATCCGCTGAAATCTCAAAACCGGTGAA 12290
 QY 1381 tcaactgcgcgttcgcgtatcctccaagaagatcaacaagatcaactcgaagcagagatg 1440
 |||||
 Db 12289 TCAACTGCGCTCCGATTCATCTCCAGAGAAATCAACAAGTCAACTCGAAGAGGAGATG 12230
 QY 1441 gtcaagaagaacatgctcgtcgtcgttaaaccaaatctcgttacttgcttgaacgcag 1500
 |||||
 Db 12229 GTCMAAGAAACATGCTCGCGCTAAACCAATTTGCTTACTTGGCTTACGCGCGCG 12170
 QY 1501 tggccaaagtcacaagatcgtcgttaagttgactcactactgagagaagttaaagaacat 1560
 |||||
 Db 12169 TGGCCCTAAAGTCTCAGAGATTCCGTTAAAGTTGATCTCAGTACCTGAGAAATTAAAGAACAT 12110
 QY 1561 cttaacgcgataacccgttaccgagagagacgtctcgttctcccccgcgagagagagagag 1620
 |||||
 Db 12109 CTTTACGCGGATTAACCGTTACGAGAGAGACCTCTGTTTCTCCCGGAGAGAGAGAGAG 12050
 QY 1621 gaagaagaagatacactcctcgttctcgttcaacgagagagacatgataatcgagatta 1680
 |||||
 Db 12049 GAAGAGGAAGGATACATCTCTCTGTTTCTGTTCAAGAGAGAAATCGAGATTGA 11990
 QY 1681 cagatgattaaacgcgcgtttagtagaggttgaagcaaggttaaatcttcgtcaaggtt 1740
 |||||
 Db 11989 CAGATATGTTAAAGCCGCTTAGCTTAGAGGTTGAAGCAACGTTTAACTTCCGTTCAAGGCTT 11930
 QY 1741 ccgtacgagattccaacggtatcattcaccgcgcgcgaactgatttggcgaagcaggttcgtgtga 1800
 |||||
 Db 11929 CGGTACGATTTCAACGATTCATTCATCGAGCGAGATGATTGGCGAAGCAGGTGCTGTA 11870

RESULT 4
 AY056255 2331 bp mRNA linear PLN 30-SEP-2001
 LOCUS Arabidopsis thaliana putative 9-cis-epoxycarotenoid dioxygenase
 DEFINITION (MOA2.4/AT3g14440) mRNA, complete cds.
 ACCESSION AY056255.1 GI:15810432
 VERSION AY056255.1
 KEYWORDS FLI_CDNA.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE
 AUTHORS
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 2331)
 Yamada,K., Banh,J., Banno,F., Dale,J.M., Goldsmith,A.D., Lee,J.M.,
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Yamamura,Y., Yu,G.,
 Yu,S., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
 Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
 Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,
 Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
 Shinn,P., Southwick,A., Tracy,S.E., Shinozaki,K., Davis,R.W.,
 Ecker,J.R. and Theologis,A.
 Full length cDNA of gene MOA2.4/AT3g14440 (GI:11994214)
 Unpublished
 2 (bases 1 to 2331)
 Yamada,K., Banh,J., Banno,F., Dale,J.M., Goldsmith,A.D., Lee,J.M.,
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Yamamura,Y.,
 Yu,G., Yu,S., Bowser,L., Carninci,P., Chen,H., Cheuk,R.,
 Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
 Kawai,J., Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C.,
 Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
 Satou,M., Seki,M., Shinn,P., Southwick,A., Tracy,S.E.,
 Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
 Direct Submission
 Submitted (12-SEP-2001) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
 Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
 Hayashizaki,Y. and Shinozaki,K.

COMMENT
 JOURNAL
 The Salk, Stanford, PGEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Yamada,K., Banh,J.,
 Banno,F., Dale,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S.,
 Quach,H.L., Tang,C.C., Toriumi,M., Yamamura,Y., Yu,G., Yu,S.,
 Bowser,L., Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C.,
 Koesema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M., Nguyen,M.,
 Palm,C.J., Shinn,P., Southwick,A., Tracy,S.E., Davis,R.W.,
 Ecker,J.R. and Theologis,A.

FEATURES
 source
 1..2331
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="RAFL08-11-H16 (R11094)"
 /note="This clone is in a modified pBluescript vector
 (FLC-1) as a BamHI/XhoI insert.
 ecotype: Columbia"
 1..2331
 /gene="MOA2.4/AT3g14440"
 1..122
 /gene="MOA2.4/AT3g14440"
 123..1922
 /gene="MOA2.4/AT3g14440"
 /codon_start=1
 /evidence=experimental
 /product="putative 9-cis-epoxycarotenoid dioxygenase"
 /protein_id="AA07104.1"
 /db_xref="GI:15810433"
 /translation="MASTPATAAAGRWLGNHTPOPLSSQSSDLGYCSLPMASRY
 TRKLNVSSALHTPAPLHPKQSSNSPALVYVPAKESNTKQMNLFORAAARAALAE
 FLVSEKRLHPKADPSPVOJAGNPAVNEQPVRRNLPVYCKLPDSIKGYVVRGANG
 LHEPTGHHFFPDGGMVAAYKFEHCSASYACRFQOTNRFYVEROLGRVPFPAKIGELH
 GHTGARLMLTYARAAGTVPDAHGTGVAAGLVYFNRLAMSDDLTPYQVITPNS
 DLKTVGRDFDGOLESTMLAHKVPDESGELFALSYDVNSKPYLKFEPFSPDKTSPD
 VEIQLDQPTMMHDFATENFVYVDDQVYFELPMILIGSPVYVDKNNVAFGLIDKY
 AEDSNKIWDAPDCEFTENLWAMEEPEETDVIIVIGSCMPDPSIPNESDNLKSVLS
 EIRLNKLGESTRPIISNEQOYNLWAGMVRNRMILGKTKFEATIALAEPKXSGFA
 KYDLITGKVKHLYGDNRYGSEPLFPLREGEDEBGITLTCVHDEKTKWSBLQIVNAV
 SLEVAIVKLPSRVYPGFHGTIGADDLAKOV"

misc_difference 1234
/gene="MOA2.4/AT3g14440"
/note="compared to genomic
acid sequence difference"
/replace="g"
1923..2331
3'UTR /gene="MOA2.4/AT3g14440"
misc_difference 2059 /gene="MOA2.4/AT3g14440"
misc_difference 2312 /note="MOA2.4/AT3g14440"
/note="not present in genomic sequence"
misc_difference 2315 /note="not present in genomic sequence"
misc_difference 2315 /gene="MOA2.4/AT3g14440"
/note="not present in genomic sequence"
BASE COUNT 615 a 560 c 533 g 623 t
ORIGIN

Query Match 99.9% Score 1798.4; DB 8; Length 2331;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1799; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atggctcttcaacgcaacgctgctgcttctggaagatgctgtgtgcaatcatact 60
DB 123 ATGGCTTCTTTCACGGCAACGCGTCCGTTCTGGAGATGGCTTGTGGCAATATATCT 182
QY 61 caagcccatatcgcttcttcaaaagctcgaacttgagttatgtagcttactatg 120
DB 183 CAGCGCCATATATGCTTCTCAAAAGCTCCAGCTGATGATGAGCTTACCTATG 242
QY 121 gccagctgctcaacagcgaactgaattgctatgctgcttcaacatccctcagctt 180
DB 243 GCCAGTCTGTGACACGTAAGCTCAATGTTCAATGCTGCTTCACTCTCCAGCTCTT 302
QY 181 catttcccttaagcaatcatcaaatcttccgcacatgttgttaagcccaaaagcaaga 240
DB 303 CATTTCCCTAAGCAATCATCAAACTCTCCGCATTTGTTAAACCCAAAGCCAAAGA 362
QY 241 tccaactaataacagatgaattgttccagagagcgcgcgagcgatgtgaagcgcg 300
DB 363 TCCAAACATTAACAGATGAATTTGTTCCAGAGAGCGGCGCGCAGCTTGACCGCG 422
QY 301 gaaggttccctgtcagcacagagaagctacacccgcttcttaaaagcgctgactagt 360
DB 423 GAGGTTTCTTGTACGACGACGAGAGCTACACCCGTTCTTAAGAGGCTGATCTTGT 482
QY 361 gtccagatcgccggaatattgtctccggttgatgaacagccgctccgctgatacttcg 420
DB 483 GTTCAGATCGCCGGAATTTGCTCCGGTGAATGAACAGCCCGTCCGGCGTAACTTTCG 542
QY 421 gtgttcggaataactcccgatctcatcaaaagagtgatgtgtgcacaggaactaacc 480
DB 543 GTGTCGGAAAACTCCCGATTTCCATCAAGAGATGATGTGCGCAAGCGAGCTAAACCA 602
QY 481 cttaacgagcggtgaacagcttcaacactcttcgacgagagcgatgattgacacgctc 540
DB 603 CTTACGAGCGGgtgacagctacacacttcttgcagagagagggatgattacacccgctc 662
QY 541 aaattcgaacacggttccagcttagctctgcggttactcaaacgaacccggttgtt 600
DB 663 AAATTGCAACAGGTTGAGTACGCTGCGGTTTACTCAACACTAACCGGTTTGT 722
QY 601 caggaagctcaattggtgtcgacccggttctcccaaaagccatcgtgtgacttcaaggca 660
DB 723 CAGGAACCTCAATTGGGTGACCGGTTTCCCAAAAGCCATCGAGACTTACCGGCAC 782
QY 661 accggtatgcgacactatctatctacgacagagctgacgagcggtatgtacccg 720
DB 783 ACCGGTATGGCCGACTATGCTATTACGCCAAGCTGACGCGGATTAATTCACCCG 842
QY 721 gcacaggaacggtgtagtaacgcccgttctgtctatttcaatgcccgttattgct 780

DB 843 GCACGCGGAACCGGTGATGCTAACCCCGGTTTGGTCTATTTCATAGCGCGTATTGGCT 902
QY 781 atgtcgagagatgattactcttaccgaagttcaagatctactcccaatgagatltaaacc 840
DB 903 ATGTGAGAGATGATTTACTTACCAAGTTCAGATCACTCCCAATGAGATTTAAAAAC 962
QY 841 gtgtgtcggttcgatttggatggacaatttagaatcacaatgattggccaccggaatc 900
DB 963 GTTGTGCTTGGATTTGATGAGACAAATTAGATTCACAAATGATGGCCACCGAAAGTC 1022
QY 901 gaccgggaatccggtgaactcttcgctttaaagctacgacgcttcgttcaagccttacta 960
DB 1023 GACCCGGAATCCGCTGATCTCTGCTTTAAGCTACAGAGTCTGTTCAAGGCTTACTTA 1082
QY 961 aaattcttcgattctcaacggaacgaactaaatccacggaactgagatcagctgcat 1020
DB 1083 AAATCTTCCGATTCACCGACGAGAACTAAATCACCGGACGTCGAGATTCACGCTGAT 1142
QY 1021 cagccgaagatgacagatcttcgctgattacagaagaacttcgtctgacttaccag 1080
DB 1143 CAGCCAGAGATGATCAGATTTCCGATTTACAGAACTTCGTGCTGATCTTACACAG 1202
QY 1081 caagtcgttccaagctccggaagatgataccggttggtctccggtgtgttaccagaag 1140
DB 1203 CAAGTCTTTCAACTCCCGAGATGATCCGCGTGGGTCTCCGCTGTTTACGACAG 1262
QY 1141 aacaagtcgaagaattcgggattttagaacaatacgcgcgaagatltacgaactaag 1200
DB 1263 AACAAAGTTCGCAAGATTCGGGATTTTACAAATACGCCGAAGATTCATCGMACTTAAG 1322
QY 1201 tggatttgatgcacagatcttgccttcacatctctgaagaagcttgggaagagcaga 1260
DB 1323 TGGATTGATGCTCCAGATTTGCTTCTGCTTCAATCTCTGGAAGCTTTGGAGAACGCA 1382
QY 1261 acaagatgaagtcgtgataaggttcctgtatgacttcaacaggaactcaatttcaacgag 1320
DB 1383 ACAGATGAAGTGTGATAGGCTGCTGATGATGATCCACGACTCAATTTTCAACGAG 1442
QY 1321 tctgagagaatctcaagaagtgctctgtctgaataccgctcgaactcctaaacccgtgaa 1380
DB 1443 TCTGACGGAATCTCAAGATGCTCTGTGTAATCCCGCTGAATCTCAAAACCGGTGA 1502
QY 1381 tcaactcgccgtcgatcatctccaacgaagaatcaacgaactcgaagcagaagatg 1440
DB 1503 TCAACTCGCCGCTCATCTCTCAACGAGATCAACAAGTCACTGAGAGCAGGATG 1562
QY 1441 gtcaacagaatactgctcgccgttaaaacccaattcgttacttggctttagccgagc 1500
DB 1563 GTCAACAGAAATCTCGCGCGGTAAACCAAAATTCGCTTACTTGGCTTACCGAGCG 1622
QY 1501 tggcctaaagtcgaagatctgactaagtgtatcctcaactcgtgagaagttaagaacat 1560
DB 1623 TGGCCTTAAGTCTCAGAGTTCCTTAAAGTGTCTCACTACAGGAGATTAAAGAAAT 1682
QY 1561 cttaacgagatcaacggttaacgagagagcctcgttctcccggaagagagagag 1620
DB 1683 CTTTACGGCGATTAACCGTTACGAGAGAGAGCTCTGTTTCTCCCGGAGAGAGAGAG 1742
QY 1621 gaagaacgaagatacatcctctgttctgttccagcagaagaagatactggaatccgagta 1680
DB 1743 GAAGAGCAAGATGATATCTCTGTTTCGTTCACGACGAGAAACATGGAATTCGAGTTA 1802
QY 1681 cagatagttaacgccgttagcttagaaggttgaaagcaacggtttaaactccgtcaaggtc 1740
DB 1803 CAGATAGTTAAACCGCTTACGTTAGAGGTGAAGCAACGAGTTAACTCCGCAAGGGTT 1862
QY 1741 ccgtacggaattcgaagatcaatctcgaagccgagtgatttggcgaagcaggtctgtgta 1800
DB 1863 CCGTACGAGATTTCACGATACATTCAGAGCCGATGATTTGGCAAGAGGCTGTGTA 1922

RESULT 5

[illegible]

/codon_start=1
 /evidence=not_experimental
 /product="F3f9.7"
 /protein_id="AAF1811.1"
 /db_xref="GI:8052547"
 /translation="MGKRLGRKROVREERYTPOGLTVKNKDVYKRLIRKLYESKILAP
 CYPDDSDCHDLECPICPLIYVPLINRSRCMSICECLOKNNNSAPPOKLIN
 LMGFPYPSLITPRCPFCCTPNAYEIRGVYKSKERGIQVEQYIETKIRMKKE
 MÖDDEKMQKRLSCSSSTMTGEMVSTGSELPVWPFAIYSLNLDGELASQ
 NASVNRHSRPRGNREDEVDLEMLYVMAIMLSVQETQONSASGETTSRQVY
 DNHSYVPEVPYPTIVPEAPTPSSSGSGLSCAIALERQVWGSSSHNNHNVANSY
 SMLGNCDSYDIEQEDGIDNHHHNRHHYEMQETGSNSYVSSYMTGGGFNHPFP
 PLVYPSPEFQOMMAVASMAEYHATATCAPEYTWQ"
 complement(join(22808..23063,23118..23244,23511..23706))
 /note="Hypothetical protein"
 /codon_start=1
 /evidence=not_experimental
 /product="F3f9.8"
 /protein_id="AAF1810.1"
 /db_xref="GI:8052546"
 /translation="MSGRGKGVSEPMKVEINTQYVETDARSERTVQDELTKNAIVA
 AGPDSASAEFGRCYDGGSKIGPTTNVPEFKVIMSYLSVPMSPNKRRAVSPHMLFK
 RTMKPQYONKRCCEYTKALKSTRLSQASQSGIMGVEVATQYQDLGTFPSIFI
 VLYMPEPSCIMELTRILISLKLIVSEKY"
 join(26752..26907,27020..27871,27923..28153)
 /note="Similar to expolygalacturonase precursor
 sp|000293|PGLX_ASP7u"

```

/evidence-not experimental
/product="p3f9.9"
/protein_id="AAFI1796.1"
/db_xref="GI:8052532"
/translation="MSVPSISGEFLIFIAVAASSISAPASALVGRKVDYRSGVGGAG
DKGDTWMAFLRKAMDACOMKGLPRVYVPECTFLGNAFLGPCSKRSISPIKTTLLA
PKDPALIKDSMTIRVYDYLTVSGGILIDGQSTSWPLNCRQTHNCRALPMNMRFQ
FVRESRLRISINSKMGHLNFEVSQHEDLITRIKAGDSPTNDGIKIGSSNKHJH
HYDITLGDCCALISGTFENDLNKNCPCPGISVSIGSKRGKGSVOGLVRSISFN
GTSNVRITKDDPSPPEPLVSNPLFKNLMDVIDVSPNIDJDRNCPNCSQVSLIR
NKSPSKIDIRVYKTONINGISTAKEAVKLCQSKNPPCANVDLFNINIVHRRDGPATS
VCENYGMIGKISPPSCIR"
33566..33539

```

Query Match	47.9%	Score 862.8;	DB 8;	Length 95769;
Best Local Similarity	73.1%;	Pred. No. 2.1e-245;		
Matches 1139;	Conservative	0;	Mismatches 407;	Indels 12; Gaps 2

OY	245	acccctaaacagatgatattgtttccacgaagacgagcgcgacgacgttggacacggcgagag	304
Db	33987	AGAAATCCAAAATTAATTAACCCGGGTGCGAGAAACCGCGGTATGGTTTATGACCGGGGAGAA	34046
OY	305	gtttctctgtccacgcagag--aaagctaacccgcttcctaaacgcgtactactagtg	361
Db	34047	ACGCCATGATCTCACACGACGCCCCCGCTCATCCCATCTCTAAACGGGGGTACTCTGCCG	34106
OY	362	ttcagatcgcgcgaataattgtgtccgcgtgaatgaacagccgcgtccgcgtactcttcg	421
Db	34107	TCCAAATATGCGCGGAACCTTCTCCCGGTACCGGAGAAACCTTCGTGCATAACCTTCCGG	34166
OY	422	tgttcggaanaattcccgattccatccataaagagatlatgtgcgcaacgagactaacccac	481

D	b	34167	TCACGTGGACCACTACCGAGATATCCATTCAAGGAGTTTACGTCAGAAACGGAGCAAAATCCAC	34226
Q	y	482	ttcaacgagccggtgacaaggttcaacacattcttcgacggaagcagttatggttcaacccgttca	541
D	b	34227	TTTCACAAACCAAGTCCTCCGCCACACATTTGTTGACGAGACGGTATGTTACACGCTGTCC	34286
Q	y	542	aattcgacaacaggttcacgtctagctctacgctctccggtttactcaagtaacccggtttgttc	601
D	b	34287	GGTTGGATTAACGGCTCGGTTACTACGCTGTGCTGGTTTACCGAATAACAAACCGGTTGGTTC	34348
Q	y	602	aggaaagtaaatcttggttcgacccggtttcccaaaagccatctgtatggtctacgagccaca	661
D	b	34347	AAGAACTAAATGTGTCTCGGTTTCCCAAAACCAATTGGAGAGCTTCCAGGCATT	34406
Q	y	662	ccggttatcccgacactcaagctatcttcaacgacagctcgaacccggtatagtcgacccggt	721
D	b	34407	TAGGAATGCCCAAGGCTTATGCTTCTTCATATACCGGGGGCTATTGTGGTTATGTGCAGACCGA	34466
Q	y	722	cacaacggaacccggtctagctaaaccccggttttggtctattcaatbgccggttattggtcta	781
D	b	34467	CCGGAGAGATCCGCTGTCCCTAACGCCGGCTTTGTTATTTCATAGGTCACTCTTATGACA	34522
Q	y	782	tctcgagaagatgatcttactcttccaaagtcagatacctccaatggaagtttaaaacccg	841
D	b	34527	TCCTCCAAACCAATTTACCGTACCAATGTAAGTTACTCAAAACCGGAGATTTAGAAACTT	34586
Q	y	842	tctgctcggtctgattttgtatgacaattagaatccaatgtatttgccaccggaagctcg	901
D	b	34587	CGGGTCGTACGATTTTCGACGGTCAGTTAAATCAACATATGATAGCCACCCGAAATATG	34646
Q	y	902	acccggaatccggtgaaactcttcgctttaaagctacgacgctgtttcaaaagccttactaa	961
D	b	34647	ATCCGGAACCCGAGAACTTTCGCTTAAAGTACGACGCTCGTTTCAAGACCTTACTTAA	34708
Q	y	962	aattacttcgattcttcaacccggaacactcaataacccggaagctcgagattcaagcttgatc	1021
D	b	34707	AATACTTCAGATTCAATCCATCCGACGGGAGAAATCACCAACGTCGAGATTCCGTTGGATC	34766
Q	y	1022	agccacagaatgacgacgatttcgagattcaagaagaactctgctcgtctacactcgaagc	1081
D	b	34767	AACCAACGATGATCCACGATTTCCGCGATCTACGTAGAAATTTCTGATGATTTCCAGACCAAC	34822
Q	y	1082	aagtcgctttcaagctcgcggaagatgataccggttggtgtctccggttggtttacgacaaga	1141
D	b	34827	AAGTGTATTTCGGATTGCGCGGAGATGATTAAGAGGTGTCTCCGGTGGTTTAGACAGAGA	34886
Q	y	1142	acaaggtcgcaagatccggattttaagacaataacgcgaagatccaatcgaaatcaatgaat	1201
D	b	34887	ACAAAGAAATCAAGATTCGGGATTTTGAAATAAAAACCGAAGATGCTTGCTGATTCAT	34948
Q	y	1202	ggattgatctcagaattctctctctctcttcatactctcgaagcgttgggaagagccgga	1261
D	b	34947	GGATTCAAATACCAAGTATTTTTTTTGTTCATCTATAGGAATCTTGGGAAGAACCGAA	35006
Q	y	1262	cagatbaagctgctcgatgaaggtccctgatctacctccacacactcaatttccaacgagt	1321
D	b	35007	CAGACAGGTTGTGTCGATTTGGATCTATGATAGCGCACTGATTAATTTTCAAGCAAC	35066
Q	y	1322	ctgacgagaatctcaagagtgtctctgtctgaatccgcctgaatctcaaacccggtgaa	1381
D	b	35067	ACGACGAATACACTTCAGAGTGTGTTTGTGCGAGATAGGCTTAAACCTTAAACACGCGGAAT	35122
Q	y	1382	caacttcgccttcgatacatcttccaagaagatccaagaatccaactcgaagcaggaatg	1441
D	b	35127	CAACACGTAAGACCGGTTATCTC-----GGAAACAATTTAATCTCGAAGCTGGTATAG	35177
Q	y	1442	tcaacggaacatgctcgtgcgttcaaaccaaatctgcttactctggtttagccgaagcgt	1501
D	b	35178	TCAACCGGAATTTATTTAGTTAGTAAGAAAAACCCGGATGCTTATCTGCTTTAACCGAACGCT	35237
Q	y	1502	ggcctaagaatctcaggaatcgcgttaaagtgtatctactactcgaaggaagttaagaaacatc	1561

Df 35238 GGCCTAAAGTCCGGTTCCGTAAAGTGACTTATCGACCGGAGATTCGAAGTATA 35297
Qf 1562 tttaacgagataacggttaacgagagagagcctctgttcccccagagaagagagag 1621
Df 35298 TTTACGAGACAGGGGAATATCGAGAGAGACCTCTGTTTACCTTCCGGTGACGGAAG 35357
Qf 1622 aagaagaagataacatcctctgttctcaacgagagaacatggaatcggagttac 1681
Df 35358 AAGACGAGATTACATATATGTTGTTCTCAGCAGAGAGAAAGTGAATCGAATCC 35417
Qf 1682 agatagtaacgcttagcttagaggttggaagcaacggtttaacttcgtaaggttc 1741
Df 35418 AACTCTAAACGCCCTTAATATGAACTTGAACCTTACCGTTCGTCGAGAGTGC 35477
Qf 1742 ctaacgattcaacggttaacatcgaagccgattgagcgaagcaggtcgtgtg 1799
Df 35478 CTTACGCTTCCACGAGAGTTCATCATAGAGAAATTTATCGAAGCAAGCTTTGTG 35535

RESULT 6
AX148316 AX148316 1818 bp DNA Linear PAT 08-JUN-2001
LOCUS Sequence 15 from Patent EP1116794.
ACCESSION AX148316
VERSION AX148316.1 GI:14347203
KEYWORDS
SOURCE .
ORGANISM tomato.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanales; Solanales; Solanum;
Lycopersicon.

REFERENCE 1 (bases 1 to 1818)
AUTHORS Iuchi, S., Kobayashi, M. and Shinozaki, K.
TITLE Transgenic plants carrying neomycin cleavage enzyme gene
JOURNAL Patent: EP 1116794-A 15 18-JUL-2001;
Riken (JP)

FEATURES
source Location/Qualifiers
1..1818
/organism="Lycopersicon esculentum"
/db_xref="taxon:4081"
1..1818
CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC4119.1"
/db_xref="GI:14347204"

translation="MATTSHTATNTWIKTKLSMPSEKRGFASNSISILIKNOHROSL
NINSLQAPRIIHPFKOSSNYOTPKNNITISHPKQENNNSSSTSKMNIYKRAAMAT
DAVESALTRHELEHPLKPTADRVQISGFAPVNPVQSLPYTGKIPKCVQGVYVR
NGANLPEPTAGHLEFDGDGAVHVOFKNGSASVACRFETERLVOEKALGRPVPKA
IGELHSGISIALMLFVARGLEGLVDHSGVANAAGLVFNNRLLAMSEDDLPHYK
VPTGDLKTEGFEDGOLKSTMIADPKIDPVSGELFALSYVIOIKPYLKYRFSGNG
EKSNDVEIPEDEPTMMHDPATENFVVIDOOVFRMSMIRGSGPVYDKKVRSG
ILDKRAGSGLDKWYEPDRCFPHLMNMEAEETDEIYVIGSCMPPOSIFNECEGL
KSVSEIRLNLKTKSTKSTIENPDEOVNPRNKRKLURKEIYALIAEWPVK
VSGFAKVLFTGEVEKFLYGDKKYGEPELFLPRDNGNSEEDDGYLLAEVHDEKWKSE
LOIVANMSIKLEATVAKLPSRVYGFHGFINDLANOA"

BASE COUNT 569 a 383 c 369 g 497 t
ORIGIN

Query Match 45.8%; Score 824.6; DB 6; Length 1818;
Best Local Similarity 69.6%; Pred. No. 2,5e-234;
Matches 1134; Conservative 0; Mismatches 489; Indels 6; Gaps 1;

Qf 170 ctccagctcttcaattcccttaagcaatcatcaactcctccgcacattgttgaagccca 229
Df 185 CTTCAATTTATCAACACCAAGAAATATATCAATTTTACACCCAAACAAAGAAACACA 244
Qf 230 aagccaaagaatcccaactaaacagatgaattgttcagagagcgcgagcgagcgt 289
Df 245 ACTCCCTCTTCTTCTTCAACTTCCAAAGTGAATTTAGTGCAGAAAGACGACCAATGGCTT 304

Qf 290 tggacgcgagcgaggtttctctgttcagccaagagaagctacacccgcctctcctaaacgg 349
Df 305 TAGATCTGTAGAAATGCTTTTAACTTAACATGAACCTTGAACACCCCTTGGCGAAACAG 364
Qf 350 ctgactcctgtgttcagatcgcgcgaatttgcctcgggtgaatgaacacccgctgcgc 409
Df 365 CCGACCCAGAGAGCCGAAATTTCTGGGAAATTTGCTCCGGTACCGGAAATTCAGTCTGTC 424
Qf 410 gtaatttcggtgtgcgcgaataactccgattcctcaatcaagagatgtgtgcgaag 469
Df 425 AATCTTTCGGGTCCACCGGAAATATCCCAATGTGTTTAAGGCGTTTACGTTTCGAAGC 484
Qf 470 gagctaacccacttccagcggtgacaggttaccactctcgcagagagcgtatgg 529
Df 485 GACCTAACCCCTTTTGAACCAACCGCGACACACCATTTCTTTCGACGCGAGAGTATGG 544
Qf 530 ttaacgcgttcaaatcgcgaacggttcagctagctagcgttcgcggttactagacta 589
Df 545 TTTACGCGCTTCAATTCAAAATAAGGCGCTAGTTACGCTTGCCTTTCACGAAACAG 604
Qf 590 accggttctgtagaagcgttcaattggtgcgcacggttttccccaagacatcggtgagc 649
Df 605 AAGAGCTTGTTCAGAAAGAAAGCTTTGGGTGCGCCCTGTTTCCCTTAAGCCATTTGTGAAT 664
Qf 650 ttaacgycacacacggttatgcccagctatcttctacgccaagcgtgcagccggtga 709
Df 665 TACATGCTACCTCTGGAATTTGCAAGGCTTATGCTGTTTACGCTCGTGCGCTTTCGAC 724
Qf 710 tagtgaccgc 769
Df 725 TGTGTTATCACAGTAAAGAAAGCTGTGTTGCAAAACCCGCTTAAATGCTATTCATAAAC 784
Qf 770 ggtattgctatgtcgc 829
Df 785 GATTACTTGTCTATGCTGAAGATGATTTGCTTACCATGTAAGTAAACACCCACCGCGG 844
Qf 830 atttaaaacgcgttgc 889
Df 845 ATCTTAAACACGAGGCTCATGCTGATTTGACGCGCAGCAAGTAAATCACCATTTAGTC 904
Qf 890 accgcgaagtcgacccgc 949
Df 905 ACCCAAGCTCGACCCAGTTCCGTTGAGCTATTTCTTACATGATGATGATTCACAG 964
Qf 950 agccttaacaaatactcgc 1009
Df 965 AGCATATCCTCAAGTACTTCAAGATTTTCAAAAATGCGGAAATCAATATGATGTTGAA 1024
Qf 1010 ttaacgttgcacgc 1069
Df 1025 TTTCAAGTTGAAGACCCCAATATGATGATTTCCCAATTACTGAGAACTTCTGCTCA 1084
Qf 1070 tacttaccagcagcgcgtttcaacgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1129
Df 1085 TTTCTATCAACAGTCTGTTTCAAGATGCTGAATGATCCTGAGGTTTACCCGCTG 1144
Qf 1130 tttacgacaagaacaaggttcgc 1189
Df 1145 TTTACGACAAAGAACAAAGTTTCCCGATTTGGTATTTCTGATAGTACGGGAAGATGG 1204
Qf 1190 cgaacatlaagtgatgtatgtctcagatgtctcgtctcgtctcgtctcgtctcgtctc 1249
Df 1205 CTGATTTGAATATGGTTGAAGTACCTGATGTTTCTGTTTCCACCTGGAATGCTGGG 1264
Qf 1250 aagagccagaagaagatgaagtcgttgatggggtccggttgcgttgcacacagatcca 1309
Df 1265 AAGAAGCAGAAACAGATGAATGTTGTATTTGTTTATGATGACACACGAGTCTCA 1324
Qf 1310 ttttaacgagtcgc 1369
Df 1325 TTTTCAATGATGATGATGAAGGCTTAAAGAGTCTTTTATTCGGAATCCGCTCAATTTGA 1384
Qf 1370 aaacgcgtgaatcaactcgc 1429

Db	Accession	Source	Length
Dn	1385	AACAGGGAATTCACAAGAAATCCAFATATCGAAAACCAGATCAACAATGGATTAG	1444
Oy	1430	aagcaggatgttcacaagaacatgctcgccgctaaccacaaattcgcttacttgctt	1489
Dn	1445	AAGCTGAATGTGTGAACCGAATAACTCGAAGGAAAAAGAGTATGCTTATTGGCTA	1504
Oy	1490	tagecgcgctgtgcctaaagtctcaaggatcgcttaaatgtgatctcaactctgaaga	1549
Dn	1505	TGCCTGAACCATGGCCAAAGATTTCTGGTTTTGGCAAAAGTAACCTGTTCACCCGGTAAG	1564
Oy	1550	ttagaaacactcttacggcgataacgcttacggagagagccctcgtttccccgag	1609
Dn	1555	TTGAGAATTTATTTATGTGTGACAAATATGGGGGAACCTTTTTCACCAAGAG	1624
Oy	1610	-----aaggaggagagaaacgaagatcatcatcctctgttgttcacagagaaga	1663
Dn	1625	ACCCAACAGCAGAGAAAGAACAGCATGGTTATTTATTTAGCTTTCCTACCATAGGAAG	1684
Oy	1664	cattgaatatcggaattacagatagttlaaacgccgttagcttgaagttgaagcaagctta	1723
Dn	1685	AATGGAATTCAGAACTGCCAATTTGTTTAAACGCAATGAGTTTGAAAGTTGAGAGCAACTGTGA	1744
Oy	1724	aacttcgtaagggttcgcgaagattcacgattcacgtaacctatcgtgaacgcgatgttg	1783
Dn	1745	AGCTTCATCAAGAGATTCCTTATGGATTTCANAGAACATTCATTAACGCCAATGATTTGG	1804
Oy	1784	cgaagcag 1792	
Dn	1805	CMAATCACG 1813	
RESULT 7			
LENEOXANT LOCUS DEFINITION	2171 bp mRNA linear	PLN 22-MAY-1998	
LYCOPERSICON ESCULENTUM mRNA FOR NINE-CIS-EPOXYCAROTENOID DIOXYGENASE.			
VERSION 297215	G1:2769641		
KEYWORDS nine-cis-epoxycarotenoid dioxygenase.			
SOURCE tomato.			
ORGANISM Lycopersicon esculentum			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
Assteridae; euasterids I; Solanales; Solanaceae; Solanum;			
Lycopersicon.			
REFERENCE 1 (bases 1 to 2171)			
AUTHORS Burridge,A., Griever,T.M., Jackson,A., Thompson,A. and Taylor,I.B.			
TITLE Structure and expression of a cDNA encoding a putative neoxanthin			
JOURNAL cleavage enzyme (NCE) isolated from a wilt-related tomato			
AUTHORS (Lycopersicon esculentum Mill.) library			
TITLE J. Exp. Bot. 47, 2111-2112 (1997)			
JOURNALS 2 (bases 1 to 2171)			
Burridge,A.			
REMARK Direct Submission			
REFERENCE Submitted (01-JUL-1997) Burridge A., The University of Nottingham,			
AUTHORS Physiology and Environmental Science, Sutton Bonington Campus,			
TITLE Loughborough, Leicestershire, LE12 5RD, UK			
REMARK Revised by [3]			
REFERENCE 3 (bases 1 to 2171)			
Burridge,A.			
REMARK Direct Submission			
REFERENCE Submitted (09-JAN-1998) Burridge A., The University of Nottingham,			
AUTHORS Physiology and Environmental Science, Sutton Bonington Campus,			
TITLE Loughborough, Leicestershire, LE12 5RD, UK			
COMMENT On Jan 13, 1998 this sequence version replaced gi:2243153.			
FEATURES			
Source			
CDS			
/organism="Lycopersicon esculentum"			
/db_xref="taxon:4081"			
79..1896			
/codon_start=1			
/product="nine-cis-epoxycarotenoid dioxygenase"			

[illegible]

Db	758	GGCTATGCTGTTTACGCTGCGGCTCTTCGACTTATGATACAGATAGAGAACTG	817
Qy	734	gtgtagctaaagccggttggctatctcaatgcccgttattgctctgtccgagatg	793
Db	818	GAGTTGGCAAGCGCGTTAGTCTATTTCAATTAACCGCATCTTCTGTGTAAGATG	877
Qy	794	attaccctaccagttacagatcccaatggagatttaaaacccgttgctgctcg	853
Db	878	ATTTGCTTACCATGTAAAGGTAAACCCACCGCGGATCTTAAACAGAGGCTCATTCG	937
Qy	854	atttgaatggaacaattagaaatccacaatgatgtccaccaccgaaagtcgaccggaatccg	913
Db	938	ATTTCGACGGCGACGCTAAATCCACCATGATAGCTCACCCAAAGCTCGACCCAGTTCCG	997
Qy	914	gtgaactctcgctttaaagctgaagcgtcgttcaaaagccttaactaaatactcgat	973
Db	998	GTGACCTATTGCTCTTGCTGACGATGTGATTCAGAACCCATACCTCAAGTATTTCAGAT	1057
Qy	974	tctcacccgagcgaactaaatccaccgagctcgagatccagttgatacagccaagatga	1033
Db	1058	TTTCAAAAAATGGGAAAAATCAATGATGTGAAATCCCACTTAAGAACCCAAATGA	1117
Qy	1034	tgacagatctcgagatcacaagaaatctgcgtcgtacacctgacccaagaagtcgtttca	1093
Db	1118	TGCATGATTTGCAATTAAGTACGAAATTCGCATCATTCGTGATCAACAGTCGTTTCA	1177
Qy	1094	agctgcggaatgatcccggtgtgtccggtgtttaaagcaagaagaagtcgcaaa	1153
Db	1178	AGATGCTGAAATGATCGTGGAGGTTCACCGTGTTTACACAAAGAAATGTCCTCC	1237
Qy	1154	gattcggatatttgaacaataacgcccgaagatctcgaacataagtcgagatgtctc	1213
Db	1238	GATTGGGATTCGTGATTAAGTACGCGAAAGATGGGCTGATTTGAATGGTTGAAGTTC	1297
Qy	1214	cagatgtctcgtctcatctctcgtgaagcgtctggaaagccagaacagaatgaagtcg	1273
Db	1298	CTGATTTGTTCTGTTCCATCTCTGGAATGCTTGGGAAGAACACAGAAATGAAATCG	1357
Qy	1274	tcgtgatagggctcgtgttgactccaacagatcaatttcaagaagtcgacgaatc	1333
Db	1358	TTGTAAATGGTTTCATGATGATGACACCCAGATCCATTTTCATTAATGTGATGAGGAC	1417
Qy	1334	tcaagagtcctcgtctcctaaatccgctgaatctcaaaacccggtgaatcctgcgctc	1393
Db	1418	TAAAAAGCGTTTATCCGAATTCGTCATTTAAAAACAGGGAATACACAAGAAAAG	1477
Qy	1394	cgatcatctccaagaaatcaacaagatcaactcgaagcagagatgtgtcaacagaaca	1453
Db	1478	CCATTAATGAAACCCGGATGAAACAGTAATTTAGAGGCTGGAATGGTGAACAGAAACA	1537
Qy	1454	tgctgcgcgttaaaacccaatcgttactgttgctttagcgcgagccggtgcttaagctc	1513
Db	1538	AACCTTGAAGGAAACACAGTATGCTTATTTGGCTATTCGTAACCATGCGCCAAAATTT	1597
Qy	1514	cagagatcgtcaaaatgtatctcactactggaaggtttaaagaacatctttagcgagata	1573
Db	1598	CTGGTTTGGCAAAAGTAAACCTGTTTACAGGTGAAGATTGAGAAATTCATTTATGGTGACA	1657
Qy	1574	accgltacggagagagcctcgttctcccccggag-----aaggaagaggaagaagc	1627
Db	1658	ACAAATATGTTGGGAGGCTCTTTTAAACCAAGAGACCCAAACAGCAAGGAGGAAGCG	1717
Qy	1628	aaggaataactcctcgttctcgtacgcagcagaagaacatgtaaatccgagttacagatg	1687
Db	1718	ATGGTTATATTTTACCTTTCGTTTCATGATGAGAAAGAAATGCAATCAGAGCTTCAATTTG	1777
Qy	1688	ttaagccgttaagcttaagagttgaagcaacggtttaaacttcgccaaggttcgctagc	1747
Db	1778	TTAAATGCAATGACTTTAAAGTTAGAGGCAACAGTGAACCTTCATCAGAGAGTCTTATG	1837
Qy	1748	gatttcaagttacatctcagcagatgattgttggcgaagcagc 1792	

Db	1838	GTTTTCATGAACTTTCATTAATGCAATGATTTGGCAATCAGC	1882
RESULT	9		
LOCUS	AX148312	1839 bp	DNA
DEFINITION	Sequence 11 from Patent EP116794.	linear	PAT 08-JUN-2001
ACCESSION	AX148312		
VERSION	AX148312.1	GI:14347199	
KEYWORDS			
SOURCE			
ORGANISM			
	Vigna unguiculata		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;		
	Vigna.		
REFERENCE	1 (bases 1 to 1839)		
AUTHORS	Iuchi, S., Kobayashi, M. and Shinozaki, K.		
TITLE	Transgenic plants carrying neoxanthin cleavage enzyme gene		
JOURNAL	Patent: EP 116794-A 11 18-JUL-2001;		
FEATURES			
source	Location/Qualifiers		
	1..1839		
	/organism="Vigna unguiculata"		
	/db_xref="taxon:3917"		
	1..1839		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC41197.1"		
	/db_xref="GI:14347200"		
	/translation="MPSASNTWPNATLPSPPFDLSTSSPTNLLPLRKSSNTIT		
	CSLOTLHPKROYPTSTSTATPTTTPPTKTTTITTTTTPRETNPLSDTQPLPKM		
	NFLQKAAATADIVETALVSHERRHRLPKDTPVOLAGNAPVPEHAADGLGVK		
	IPKCTIDYVYVNGANPLPEYAGHHFPEGDDMVNAKFTNQAASYACRTFTQLSOE		
	KSIDRPVFPKAIIGELHSGIARLLLEFARLFDLVDSQGMVANGLVFNHLLA		
	MSEDDLPYHVDITPNNGDLTVGRVDFNQLSTMIAPKLDVPVGDJHALSYDIQK		
	YLKFRSPDVKSPDVEIPLKEPTMHMDFATFNFLVVPDQVYFLEMTIGSPV		
	VYDNKTSRFGLILKKNKADANAMWIDAPDCEFLMVAWEPEVEVYVIGSCMTPA		
	DSINKECESIKSVLSERLNLPTGKSTRRG11SDAEQVNLAEAGVNRNKGRTQFA		
	YLAIAEPKVSIGFPAKVDLSGEVKKHYGGEKRGGEPLFPNOKEDDGIILAFVHD		
	EKEWKSLEQIVNAONLKLKIASIKLPSRPVIFGHTFHSKDLRQA"		
BASE COUNT	447 a	595 c	442 g
ORIGIN			355 t
Query Match	44.4%	Score 798.4:	DB 6; Length 1839;
Best Local Similarity	70.4%;	Pred. No. 1.6e-226;	
Matches 1098;	Conservative 0;	Mismatches 456;	Indels 6; Gaps 2;
Qy	233	ccaagaatccaacaactaaacagatgaatttggctcaagagcggcgcgagcgttgg	292
Db	281	CCAACCAACCATTAACCTCAAAAATGGAACCTTCTCAGAAACCGCTGCCAGCCTTGG	340
Qy	293	acgcggcggaggggttctctgtcagccagcagaagctaacaccgcttctaaacgctg	352
Db	341	ACCTGCTCGAAGCGCGCTCGTGCAGCAGCGCAACACCCGCTCCCAAAACGGCGG	400
Qy	353	atcctagtgcttaagatcgccggaatttgcctcgggtgaatgaagaagccgctcggcgt	412
Db	401	ACCCAGAGGCTCCAAATCCGCGGGAACCTTCGCGCGTCCGAGCATGCGCGGATCAAG	460
Qy	413	atcttcgagtggtggaaactcccgatctcaataaagagttgatgtgccaacggag	472
Db	461	GACCTCCCGGTGTCGGAATAATCCCAATGATGACGCGCTGTAACGTGCGCAACGGTG	520
Qy	473	ctaaccacttacagcggcgtgacaggttacaccacttctcgaaggaagcgtatgcttc	532
Db	521	CCAAATCCGCTTACAGAGCCTGTGGCCGACCACTTCTTCAGACGGCAGCATGCTCC	580
Qy	533	acgcgcgtcaaatcgaacacggtttagctagctacgcttgcggtttactaagactaac	592
Db	581	ACGCGGTGAAGTTCAAGAACGCGCGCGCAGCTACGCTCGCGCTTCAACAGAGCAGCAGC	640

QY	593	ggttgttcaggaacgcftcaatttggttcagacggttttccccaagaacatcggttagcttc	652
Db	641	gctcttcgcaagaacaaatctcttaagcccccgggtgtttccgaaggccatcgaggagcttc	700
QY	653	acggcacaaccggtatgcccagctcatgtcatctcaagccagagctctgaacggtatag	712
Db	701	acggcacttcggcgatcgccgggtcctctcttcttaccggccggtctcttcggggctcg	760
QY	713	tcgacccggcacacggaacccggtgtagctcaacgcccgtttgttctatttcaatgycggt	772
Db	761	ttgatgggtccccaaggcatggcgctggcgaacccggctcgcgtacttcaacacaccc	820
QY	773	tattggtcatgtcggagatgatattactactccaagtccaagtccaactcccaatgagatt	832
Db	821	tcctggcccatgtccgaagacgattttacccttaccctgagatcacccttaacggcgact	880
QY	833	taaaacccgttggtcggttcgatttgatttgatgagcaattagaatccaacagatggccac	892
Db	881	ttaccaccggttgccgctttagactttcaacggcgactcaactcaacatgatcgccac	940
QY	893	cgaagctgcacccggaaatccggtgaatctcgtcttlaagcttcaagctcgtttcaaacg	952
Db	941	cgaaactggaccctcgtagagggacctccagcgcgtacagctacacgctattcgaagac	1000
QY	953	cttactcaaatactcccgattctcacacggaggaactaatcaeccgagctcgagattc	1012
Db	1001	cttaccttcaacttaccttccgttttcccccccgccctcaagctcccccgactggaatcc	1060
QY	1013	agctgtacagccaagatgatagcatttcgcgttcaagagaaactcgccgctac	1072
Db	1061	ccctgaagaccctcacatgatgatgacatttgcctaaacgagaaatttgcgtcgtcc	1120
QY	1073	ctgacagcaagctcgtttcaagctgcgcggagatgaatccggtgtgctccggtgtt	1132
Db	1121	ccgaccacacagctggtcttcaactaaacgagatgatcaacccggcggtccccggtgct	1180
QY	1133	acgacaagaacaaggtcgcgaagtcgggattttagacaatacgcgcgaagatcatcga	1192
Db	1181	acgacaaacaaaacccctcacgggttgggtttctgcacaaagatgggaagacgcgaatg	1240
QY	1193	acattgaatgtgatgtgacgtccagattgcttcgtccatctcpgaacgcttggag	1252
Db	1241	cgatccggtgatgcagcgcgcggattgttcttgccttccaccttggaacggctggagc	1300
QY	1253	agccagaacacgatgaagctcgtcgttgaaggtccctgtatgctccacccaagctaat	1312
Db	1301	agccgaacacgagagaggttggtgatgggtcgtcgaacacccctgcgactccattt	1360
QY	1313	tcacaggtctgtacggaatctcaagagtgctgtctcctaactccgctgaatctcaaaa	1372
Db	1361	ttcaacgaattggaagaaagtttgaagcgtcgtctcaagatgaagcgtgaacttgagga	1420
QY	1373	ccggtgaataacatcgccgcgtccgatactctccaagagaatacaaacgtcaaccccgag	1432
Db	1421	ccggcaactccactcgcgcgcgccttatttctcgacccgc -- aacaaagaaactcggtag	1477
QY	1433	caaggatgtgcaacagaacaatgctcgcgcgttlaaaccaaatcgttactgtgttag	1492
Db	1478	ccggcatgtgtaacagaacaagctcggaaggaacccagtttcgggtatcgtgctcg	1532
QY	1493	ccgagcgttggtcctaagtctcaagatcgtctaagtgtatctcaactactgagaagta	1552
Db	1538	cgagaccccttgcccaaaagcttcggcgcttgcgaagtattgttgctgagtggggaagta	1597
QY	1553	agaacactttagggcgataccglttagggagagagctctgtttctccccggaag	1612
Db	1598	agaaattacattatggaagaaagaaagttcgttgaggagcctctgttcttccc -- aacg	1654
QY	1613	gaggagaggaaggaagatacctctgttcttcaagcagagaagatggaat	1672
Db	1655	gccaaaaaagaagacgattgggtatattcttgscattctgtcagcagcaagaaagattgaaat	1714
QY	1673	cgaggttcaagatagttacgcggttaggttgaagcaacggttlaacttcgt	1732

Db	1715	CCGAGCTCGAGATTGTGAATGCCCAAAATTTAAAGTCGACGCTTCATCAACATCCCT	1774
Qy	1733	caaggttcctcagcagattccagctacatccatcgagacgatgattgctgcagcag	1792
Db	1775	CACGCTTCCCAACAGCTTTTCATGACACTTTCATTCATTCACAGAGATTGAGGAACAAAG	1834
RESULT	10		
AB030293			
LOCUS			
DEFINITION	AB030293	2432 bp	mRNA linear PLN 01-DEC-2000
ACCESSION			
VERSION	AB030293		
KEYWORDS	AB030293.1	GI:9857289	
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
gene			
CDS			
Query Match			
Best Local Similarity			
Matches 1098:			
Conservative			
Mismatches			
Indels			
Gaps			
233	ccaagaatccacaactaaacagatgatttgcagagagcgcgcgagcgttgg	292	
406	ccaaccaacacatttacctcaaaaatggaactttctccagaaacccgctgcacgcgcttgg	465	
293	acgcgagcgagaggttctctctgtcagcagcagagaagactaacccgctctctaagcgctg	352	
466	acctggtccaaacggcgctgcttcgcacagacggcaaacaccgcgctccccaacacggcgg	525	

QY	353	atccctagctgtcagatcgcgcgaatcttgtctccgltgaaatgaacagccgcgtccgcgtla	412
Db	526	ACCCGAAGGGTCCAAATTCGCGCGGAACCTTGCGCGCGGTCCGGAGCATGCCGCGATCAAG	585
QY	413	atcttcgcgtgtgtcgcgaaatactcccgatctcaataaagagttatgtgtgcgaacgag	472
Db	586	GACATCCGGGGTGTGCGAAAAATCCCAATTCGATTAACCGCGCTGTACGGCCAAACGGGTG	645
QY	473	ctaacccaactcaacgagccggttgaaagtgtaaccaactctcttgacgagagatgtgtc	532
Db	646	CCAAATCCGCTCTACGACCCCTGTGGCCGGGACACACTCTTCTTCAACGGCAGCGCATGTCTC	705
QY	533	acgcgcgtcaaatctcgaaacacggttcagctagtacgctgtccggttctactcagactaac	592
Db	706	ACGCGGTGAAGTTACGAAAGCGGCGCCGACAGCTACGCTTCGCGCTTCAACCGAGACGACG	765
QY	593	gatttgttcaaggagacgtlcaatttgggttcgacacggttttccccaaagcatcgtgtgcttc	652
Db	766	GTCCTCTCGCAGGAGAAATCTCTAAGGCGCGCGGTGTTCGCCAAGGACCATCGGGAGGCTCC	825
QY	653	acgcgcacacacggttatgtcccgacatgtcatctacgcagagctgcaacggtatag	712
Db	826	ACGGCCACTCCGGCATCGCGCGGCTCTCTCTTACCGCGCGGTCTCTTTCGGGCTCG	885
QY	713	tcgaccccgacacacggaacccggtgtgtagctaaacgcggtttgttctatttcaatggccggt	772
Db	886	TTGATGGGTCTCCAGAGGATAGGGGTGGCGAGCGCGGTCTGTCTACTTCAACAMCACC	945
QY	773	tattgtctatgtcggagagatgtattactctaccagttcagatctcccaatgtgagat	832
Db	946	TCCTTGGCCATGTCCGAAGAGATTTACCTTACACCTGAATTCACCCCTAACGGCGACT	1005
QY	833	taaaacacgtttgttcggttcgatatttgatggaaatctagatccaaatgattgcccac	892
Db	1006	TAAACACCGTTGGGCCCTTACGACTTAAACGGGACGCTCACTCAACAAATGATCGCCACC	1065
QY	893	cgaagctcgaccccggaatcccggtgaactcttcgctttaaagctacagcgttcgttcaaac	952
Db	1066	CGAAMTGTGACCCCGGTGACGGGCGACTCCAGCGCGTCAAGTACAGACGTCAATCAGAAAGC	1125
QY	953	cttaccctaaatactctcgaatcttcacccggaacgaactaaatcacccggaacgtcgaattc	1012
Db	1126	CTTTCCTCAAGTACTCTCCGTTTCTCCCGACGCGCTCAAGTCCCGCGACGTGGAATCC	1185
QY	1013	agcttgtacgacaaagatactcaagattttcgcgagttacagagaactctcgctcgctac	1072
Db	1186	CCCTGAAAGAGCCCAACCATATATCAGAGATTTGCCCTAAGGAGATTTTGTGTGTCTC	1245
QY	1073	ctgacacagcagtgctgttcttaagctctgcggaatgataccgcggtgtgtctccggtgttc	1132
Db	1246	CCGACCGACGAGGTGTCTTCAAACTAACGGAGATGATCAACGGCGGGTCCCGCGTGTCT	1305
QY	1133	acgaacaaagaacaaagttctgcaagatctcgggaattttagaacaatacagccgaagaatlcacga	1192
Db	1306	ACGCAAGAACAAAACTCTACGGTTTGGGATCTTGCCACAAGATGCGAAGGACGCGAATG	1365
QY	1193	acattaaatgtgatgtatggtcgcagagattgtctgcgtctcatctatctggaaagctttggag	1255
Db	1366	CGATGCGGTGTGTGACGCGCCCGGATTTGTTTGTGTTCACCTCTGGAACCGGTGGGAGG	1425
QY	1253	agccagaacaaagatgaagttcgtgtatagaggttcctgtlatgactccacagaactcaattt	1312
Db	1426	AGCCCGAAGACGAGAGAGTGTGTGTATTTGGGTCTCGATGACCCCTGCGACTCTCATTT	1488
QY	1313	tcacagagttctgacgagaatctcaagagtgtccgtctcgtgaaatcccgccctgaaatctcaaa	1372
Db	1486	TCAACGATGTGAGGAGATGTTTGAAGAAGCGTCTGTCAAGATTAAGGCTGAACCTTAGAGA	1545
QY	1373	ccggttgataactacccgcgcgtcgcgatcatctccaagaagatacaaaagtcaaccttcgag	1433
Db	1546	CCGGCAAGTCCACTCTGGCGCCCATTTATCTCCACGCCG---AACAAAGTACCTGTGAAG	1602

OY	1433	caggatattgtaacaggaacatgctcggccgttaaacccaatttcgcttaacttgcttag	1492
Db	1603	CCGGCATGTGTGAACGAAACAAAGCTCGGAGAGAGACCCAGTTCGGTAATCGCTCTGG	1662
OY	1493	ccgaagccgttgagcctaagaatctcaaggattcgcctaaagttgactctactctgagaaagta	1552
Db	1663	CGGACCCCTGGGCCCAAACTTCGGGCTTTGGCAAAAGTTGATTTGCTGATGGGGAAGTA	1722
OY	1553	agaacatcctttacggcgataaccgtttacggagagagacccctctgtttccccggagag	1612
Db	1723	AGAACTACATGTATGTGGAAGAGAAAGTTCCGGGGGAGACCTCTGTTTCTTCCC---AACG	1779
OY	1613	gagagagagaaagcgaagatacatcctctctgttttcgttcaagcagagaaagataagaaat	1672
Db	1780	GCCAAAGGAAGACGATGGGTATTTCTGGCATTCGTGCACAGACAGAAAGAAATGGAAT	1839
OY	1673	cggaaattacaatagttaacgccgttgaacttagagagttgaagcaacggtttaacttcgt	1732
Db	1840	CCGACCTGCAGATTGTGTAATGCCCAAAATTTAAAGCTTGAAGCTTCATCAAACTCCCT	1899
OY	1733	caagagttccgtacaggaattcaaggtataatcaatcagagacccgaatgatttggcgaagcag	1792
Db	1900	CTCGGTTTCCCTACCGGTTTTCATGGAACTTCATTCATCAAGATTTTGGGAAACAG	1959

RESULT	11
LOCUS	AF190462
DEFINITION	Phaseolus vulgaris 9-cis-epoxycarotenoid dioxygenase (NCED1) mRNA,
ACCESSION	AF190462
VERSION	AF190462.1
KEYWORDS	GI:6715256
SOURCE	Phaseolus vulgaris.
ORGANISM	Phaseolus vulgaris Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
REFERENCE	Phaseolus. 1 (bases 1 to 2398) Qin,X. and Zeevaart,J.A. The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory step of abscisic acid biosynthesis in water-stressed bean Proc. Natl. Acad. Sci. U.S.A. 96 (26), 15354-15361 (1999)
JOURNAL	20079657 10611388
MELINE	2 (bases 1 to 2398) Qin,X. and Zeevaart,J.A.D. Direct Submission Submitted (27-Sep-1999) MSU-DOE Plant Research Laboratory, Michigan State University, Wilson Str., East Lansing, MI 48824, USA
PUBMED	Location/Qualifiers
AUTHORS	1..2398
JOURNAL	/organism="Phaseolus vulgaris" /cultivar="Top Crop" /db_xref="taxon:3885"
FEATURES	1..2398 /gene="NCED1"
SOURCE	124..1971 /gene="NCED1" /function="catalyzes the oxidative cleavage of the C-11,12 double bond of 9-cis-epoxycarotenoids (C40) to yield xanthoxin (C15) and C25-apo-aldehydes"
gene	/codon_start=1 /product="9-cis-epoxycarotenoid dioxygenase" /protein_id="AAF26356.1"
CDS	/db_xref="GI:6715257"
	/translation="MPSPASNWTWINTTLPSSCSPPFKDLASTSSPTLLPFKRSSS NTNIITSLIOTLIHKYOYSPSTSTTTTPPIKPTTTTTTPPRETKRLSDPKOPP OKNNFLDKAATGIDMETALVSIESKHPLPKRTDPVOJAGNAIPPEBAADALPV VGKIPKCIDGVTVRGNANPLIEPVGHHFEGDDMAVAVEFTNGASVACRFETORL AQESLRDVPFAKITGELHSGIARLLFYARSLFDLVGSHGMGVANGALVFNHH LAAISEDLPYHVAVITSGNDLTIVGRVDENQLNSTMIAPHKLDIPVNGDIHALSYDVV

OKPLIKYKFRSAGSDVSPDVIPLKEPTNMMHDAITFNFPVYVPOOVYKILITETNGG
SPVYVDDNCRBSGILDNKADNANMAMKIDAPBECFCCHLNANNEBPETDILYIGSGM
TPDASITNECDKSLSVSEIRLNRKTRSTRIRIISDAEYVNLAEAGMNNKIKRGRT
QFALALAEWPKPVKSGEAKVDLFEGEVYVYKGEKGGDELFTLPNDEERDGYLLAF
VHDEKEMSELOIVNANOLKLEASIKLPSKRVYVGFHGTFLSKDLRKA"

BASE COUNT 627 a 687 c 516 g 568 t

ORIGIN

[illegible]

QY	1127	tggtttaagacaagaacaaggtcgcagaatctcgggatttttagacaataatgcggaagatt	1186
Db	1307	tggttttagacagaaacaaacacccacggtttgggatttttgatnagaattccgaagacg	1366
QY	1187	cctcgacaatcaagtgaggtctatgctccagaattcttcctgcttcacatctctgaagcctt	1246
Db	1367	cgaaatgcgcatggcggtgcatgcacggccggagtggtttctgcttcacatctcgaacccgt	1426
QY	1247	gggaagaccgagaacaagatgaagtcgctcgatgataaggctcctgaatgaagctcaccagact	1306
Db	1427	gggagagaccctgaaaccacacagagattgtaggattggggtctctgcgatgacacccctccgacct	1486
QY	1307	caatttcaacgagctcgaagagaatctcagaagctgctcctctcgaagaatccgcctgatctc	1366
Db	1487	ccatttttaataatgatgcgcatgagactttgaaagacgcgttttctctgacgtanagaagctgaaact	1546
QY	1367	tcaaaaccggtgaatcaactcgcgcgtccgcatcatctccaaagaagatcaaaagaatcaacc	1426
Db	1547	tgagagacggcgacagtcgactcgccccccttattctccgacgcccg---aacaaagtgaacc	1603
QY	1427	tgcgaagcggatggtgtcaacagaacaatgcttcgcgcgcgttaaaaccaaatctgactacttg	1486
Db	1604	ttgaagccggcgctggtgaacagaaacaaacgacgaaagaaacacgacgttccttatctg	1663
QY	1487	ctttgaccgagccgctgagcctaagctcctaagatctgcctaaagttgatctcactactgag	1546
Db	1664	ctcttgccgagcgccctggccccaagatctcgggcctttgcgaagattgatttggttcagtgggg	1723
QY	1547	aagttaagaacaactcttaacgagcgttaacgcgttaccgaagagagcctctgcttccccg	1606
Db	1724	aagtcacacagatcacgtgatgtgagagaaagaaagtttcgggggacccctctgtttttttccc	1782
QY	1607	gagaaggggaggaagaacgaagatacatccctctgttctgcttcaacgagagaagacat	1666
Db	1783	--aacggcgagaaacaaagctgatggggtatattcttgcgcatcttgctgcacgacagaaagagct	1840
QY	1667	ggaaatcggagttacagatagttaaagccgcttagacttagaggttgaagcaacggttaaac	1726
Db	1841	ggaaatccgagctgcagattgtgaaatgcccaaaattttmaacctgaaagcttccatcaaac	1900
QY	1727	ttccgtaacaggttcgcaagcagattcaaggttaacatcatcagcagcagatgatctggcga	1786
Db	1901	tccctctcgcgcttcttatggttttcattgacacattttattcatcttaagagattttgacga	1960
QY	1787	agcagag 1792	
Db	1961	agcaag 1966	
RESULT 12			
AC074176/c		72058 bp	DNA linear PLN 19-JAN-2001
LOCUS			
DEFINITION		Arabidopsis thaliana chromosome 1 BAC T2H7 genomic sequence,	
ACCESSION		complete sequence.	
VERSION		AC074176	
KEYWORDS		AC074176.5 GI:12321625	
SOURCE		HTG.	
ORGANISM		thale cress.	
REFERENCE		Arabidopsis thaliana	
AUTHORS		Eumetopota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
		Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots:	
		Rosidae: eutroids II; Brassicales; Brassicaceae; Arabidopsis.	
		1 (bases 1 to 72058)	
		Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Greasy, T.H., Haas, B.J.,	
		Wu, D., Maitl, R., Rensing, C.M., Koo, H., Fujii, C.Y., Ullendack, T.R.,	
		Barnstead, M.E., Bowman, C.L., White, O., Niernman, W.C. and Fraser, C.M.	
		Arabidopsis thaliana chromosome 1 BAC T2H7 genomic sequence	
		Unpublished	
		2 (bases 1 to 72058)	
		Town, C.D. and Kaul, S.	
		Direct Submission	
		Submitted (15-JUL-2000) The Institute for Genomic Research, 9712	
		Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
3 (bases 1 to 72058)	Town, C.D. and Kaul, S.	Direct Submission	Submitted (19-JAN-2001)	The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtowne@igf.org
			On Jan 19, 2001	this sequence version replaced g1112280878. Address all correspondence to: atef@igf.org

BAC clone T2H7 is from *Arabidopsis thaliana* chromosome 1. The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including GeneScan⁺ (Chris Burge, <http://CCR-081.mit.edu/GeneScan.html>), GeneMark (Mark Borodovsky <http://genome.mbiology.gatech.edu/GeneMark/>), GlimmerA (a variant of Glimmer, see Mihaela Perlea, http://www.tigr.org/softlab/glimmerm_hm/glimmerm.html, and GeneSplicer (Mihaela Perlea and Steven Salzberg, <http://mperleab.tigr.org/>). Searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/cdb/tg1.shm1>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES	Location/Qualifiers
source	1. .72058

misc_feature

/note="28/97 nt beyond this point were not included in the submitted sequence due to an overlap with another BAC

CDS

```

/product="hypothetical protein"
/protein_id="FAG50853.1"
/db_xref="GI:1322167".

translation="MSYLKSDPVSRIHNEPOSTLSFDHFLDPSLLDILFDKVDADV
KDLQELSVRRSHSLPVPEENIVAVDVIDDDSSSDENHRSLMTAISDAGAA
GSGSALFLPFLPFGPEOMGLDILKPRSSSPDMSANVDEIGVTHSSPVOYLNK
NGEIRFLTELTPFEGDGLIKLWRAPDSTLDMITGASVSSIOGKSSNHSN
VDENGNIPESFTTNGSLKLRVYMTLSLTAASRRHYLLQPIINHKSLDVLISDDK
GAGVLEKQVDELELRLVPLASASARRVLVALLKMLWYAEVLDLPGLVKKATATV
GIIVSEKSKKEVCASWMLSDAFEEPECTVAKMLIKRRYLCENSS".

```

repeat_region

repeat_region

repeat_region

supervision

gene

CDS

gene complement(7213. .8049)
/gene="T2H7.2"
/note="predicted by genemark hm"
complement(7213. .8049)
/gene="T2H7.2"
/codon_start=1
/product="hypothetical protein"
/protein_id="ANG50854.1"
/db_xref="GI:123216248"
/translation="MTMPSTWIKNPKEPEPKKRYSSMLTIFLPIFTYIMITHVE
VLSVSQKDTKGLFEITSNLLILIIADYGSVDSESDQFGEYTVAAATMNRKDD
PLPVLYRNTKDEIKNPKDVEEPNPEEEDPVMKILCVSPKEITVRVSEKODD
DVAMEKRVKEQTLASEACTRNHNVPNPNPYGSKSKDERRRLSDIETTKRSKY
GKKKSDCSWMVYIPKEWEVYKEESEEFKLSNEELNKRVEEFIORFNQIKSCSPRVS
SF"
repeat_region complement(8757. .8780)
/rpt_family="Ar_rich"
repeat_region complement(10775. .10811)
/rpt_family="Ar_rich"
repeat_region complement(10873. .10905)
/rpt_family="Ar_rich"
repeat_region complement(11013. .11070)
/rpt_family="Ar_rich"
repeat_region complement(11798. .11818)
/rpt_family="Ar_rich"
gene complement(13002. .16330)

repeat_region

repeat_region

repeat_region

mRNA

gene

CDS

repeat_region

mRNA

galle

```
CDS
complement(join(19380..19710,19799..20403))
/gene="T2H7.4"
/codon_start=1
/product="unknown protein"
/protein_id="ANG50858.1"
/db_xref="GI:12321632"
/translation="MKLVYADKIPDELYDAQVGAHQWGLFSDRTNRSAIISDYLN
PYASKSKPMHILPEFTPMYSQDEVCNVNVAESPDDQDHEDMVVGIKFLGROLS
LCRPHLDLWNTILTFESMEVSELMYSKDORFLYLAQGYGLGSMDELNFEDEKPK
FHELYLHNLPMNPRSLMKOLDSYGRDHWVESPESEFLYMKFCEYSPENSTNCKSP
IYVYREBETKDKRKNRYTDIDGLCFISGDEFCYKASSCPGLQNSLSLHGRFL
ALNLTKRTMGIEYFPGIPKRIPIYLPWLPFSF"
complement(21599..21620)
/rpt_family="Ar_rich"
/complement(21745..21765)
/rpt_family="Ar_rich"
complement(join(<21856..22030,22534..22806,23042..23208,
23290..23520,24442..>24837))
/gene="T2H7.5"
/complement(21856..24837)
/gene="T2H7.5"
/note="similar to En Spm like transposon protein
GI:415361 from (Arabidopsis thaliana)"
complement(join(21856..22030,22534..22806,23042..23208,
23290..23520,24442..24837))
/gene="T2H7.5"
/codon_start=1
/product="En/Spm-like transposon protein, putative"
/db_xref="GI:12321633"
/translation="WSSSAIATAAALKERIRIMDVRSYVGLFSSVTVDNYSNRF
RKQDIVLIANGVARALLNIPEDAVEVTELHNSWTCYELVHVTKPMAAPPKRGMPDT

Query Match      41.7%  Score 750.8;  DB 8;  Length 72058;
Best local Similarity 68.2%  Pred. No. 5.2e-212;
Matches 1075;  Conservative 0;  Mismatches 492;  Indels 9;  Gaps 2;

QY 223 aagcccaagaacaaatccaaacaaacaaatgaatgtaattgltccagaagagcgagcg 282
DB 55440 AACCCATATATCCCGAAAAAGACACTTCCCGTTGGAACCCCTCCACAGCGCGCCTCC 55381
QY 283 gcaagcttggaagcgagcgagaggttcctctgtcagcaagagaagtaaccgcctctct 342
DB 55380 GCGGCACTGCACTTCCGCAAAACCGCATTTGTTAAGACGGCAGTCTTAAACCTTCCTCC 55321
QY 343 aaaaagcgatcatcgtgtgttcagatcgcgcgaatcttgcctcgcgttgaaagaacccc 402
DB 55320 AAAACAGTTGATCTCTCGCATGATGATCTCGGTAACCTAGCTCCGATACCGAGCAATCC 55261
QY 403 gtccgagcgaatactcttcggttggtcgaaaactcccgattccaataaagagtgatgtg 462
DB 55260 GTTAAATCATCTCTCCCTGATGAGAAAATCCCTGACTGATGACGGTGTATACCTC 55201
QY 463 cgcaagagagcttaaccattccaagcggtgagcaggtgcacactcttcgacgaagac 522
DB 55200 CGTAACGGCGCTAATCACTCTTCGAGCCAGTTTCGTGATCATCACTTATTCGACGGGTGAC 55141
QY 523 ggtatggttcaagcgatccaatccgaacaggttcagactagactgactgacggttact 582
DB 55140 GGTATGTTTACAGCCCTTAAATCACTAAGAGAGACGCAAGTTACGTGCGGTTTAAAG 55081
QY 583 cagaactaacggtttgttcaggaacgtcaatgtgtgcacggttttccccaagaacalc 642
DB 55080 GAAACCGAGAGATTGTTCAAGAGAAACAATCGGTTTCGATTTTCCCTTAAAGCTATA 55021
QY 643 ggttgagcttcaagcgcaacacggttatgcccgaactcaatgtatttcaagccagagctgca 702
DB 55020 GGTGAGCTCATAGTCTACTGTGAACGCAAGGTTGATCTATTATTTACACAGCTGGTTTA 54961
QY 703 gccggtatagtcgaaccgcgcacaggaacggtgttagctaacgccggttggtctatttc 762
DB 54960 TTCGGTTTATTAAATACAAAAACGAGACGGAGTTGCTAAGCCGCGTGTGTTTACTTTC 54901
```

```
QY 763 aatgacggttatgtgctatgtcgcgaagatgattacacttaacgaattcacatccccc 822
DB 54900 CACGACCGGTTAFTAGCTATGTGTGAAGATGATCTACCTTACCAGATTGCTGTCACTGAC 54841
QY 823 aatgagatttaaaaacggttggtcgttcgatctttagatgacaatttagaatccaaatg 882
DB 54840 AATGGGATTTTGTAGAGCCATTCGAGAGATTGATTCTGACAGGCAACTAAGCTCCCGCAATG 54781
QY 883 attgcccacccgaagtcgaacggaatccggttaactcttcgctttaaagcagagctc 942
DB 54780 ATCGCTCACCCCAAGATTATCCGTTACGAAGAGACTATTTCGCTTGAGCTACGACGTC 54721
QY 943 gttcaaaagccttaccataaatactccgattccacccgcgaagaaactaacccgcgac 1002
DB 54720 GTTAAAGAAACCGTATTGTAATACCTTAAATTCTCGCCGAGAGTGAGAAATCACCGGAC 54661
QY 1003 gtcagagattcagcttgatcaagccaagatgatgacgaatttcgcgattacagaagactc 1062
DB 54660 GTTGAATTCCTCTCCGCAAGTCGAGATGATGACGATTTCGCGATCACTGAGAAATTC 54601
QY 1063 gtcgtcgtacccgacgaagatcgcttccaaagctgcgcgagatgataccgcggtggtct 1122
DB 54600 GTTGTGATTCGCGATCAACAAAGTTGTTAAAGCTCTCCGATATGTTTCTTGGAAATCT 54541
QY 1123 ccggttggttcaagcaagaacaaggttcgaagattcgggattttagacaatacgcgaa 1182
DB 54540 CCGGTTAATATACGACGAGAGAAATTTCCCGTTTGAATTTTGCCGTGAAACGCTATA 54481
QY 1183 gattatcgaacattaaagtgatgtatgtctccagatgtctctgtcttccattctggaaac 1242
DB 54480 GATGCTCGGAATGTTGGTGTGAGTACCGGAGACTTTTGTTCATCTTTGGAAAC 54421
QY 1243 gcttggaagagccagaacaagatgaagtcgctcgtatagaggttcctgtatgactcaaca 1302
DB 54420 GCTTGGGATCAACCGGAGACAGACAGAGTTGATGATGATGATGATGATGATGATGATGAT 54361
QY 1303 gactcaatttcaacagatctcgaagaaactcaagagtgctcgtcgtcgtcgtcgtcgt 1362
DB 54360 GATTCGATCTTCAACGACGATGCGATGAGCTACGATCAACAGTGTTTTATCAAGATCCGGTTA 54301
QY 1363 aatctcaaacccggtgaatacaactccgctccgcatcatctccaagaagatacaaaagtc 1422
DB 54300 AATCTCAAAACCGGGAAATCCACGACGACGATATATATTCCTCGGTTGCTT--CAGATG 54244
QY 1423 aacctcgaagcaggaatgttcaacagaacaatcgtcgcgcgtgaacaaacaaatcgtctac 1482
DB 54243 AATCTGAAGCTGATATGATGTAACGAAATCTTCGCGAGAGAAACCGGTAACGATATC 54184
QY 1483 ttggtcttagcgcgagcgttggtcctaaagtctcaagattcgtcaagttgatactact 1542
DB 54183 CTAGCCATAGCTGAACCGGTGCTTAAAGTCTCAGGATTCGTAAGATTGATCTTCTTAC 54124
QY 1543 gtagaagttaaagaatactcttaagcgataccggtttagggagagagctctgtttctc 1602
DB 54123 GGAAGATGAATAATCACTTTTACGGCGGTAAATAATAGGTTGTAACCTTTTCTTAA 54064
QY 1603 cccggttcaaggaagaggaagaagaagaatacactcctctgttcttcgacagac 1656
DB 54063 CCTAGAGGTTTGAATCTAGACGAGAGATGACGATTAATATATGATGATGATGATGATGATGAT 54004
QY 1657 gagaagacatggaatccggaatcagatagatlaaacgcggttagacttagaggttgaagca 1716
DB 54003 GAGGAGAGTTGGGAATCGGAGCTTCAATATTGTAACGCCGTTACGTTGAACTAAGACA 53944
QY 1717 acggttaaacctccgtcaagaggttcgtcagatttcaaggttaacattcaatcggaacgat 1776
DB 53943 ACCGTTAAATTCGCGCAAGAGATACGATGATGATGATGATGATGATGATGATGATGATGAT 53884
QY 1777 gatttgcgaagcag 1792
DB 53883 GATATGTTAAACCAAG 53866
```

RESULT	13				
LAI48302	AXI48302	1752 bp	DNA	linear	PAT 08-JUN-2001
LOCUS	Sequence 1 from Patent EP116794.				
DEFINITION	AXI48302				
ACCESSION	AXI48302.1	GI:14347189			
VERSION					
KEYWORDS	thale cress.				
SOURCE	Arabidopsis thaliana				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (bases 1 to 1752)				
AUTHORS	Iuchi,S., Kobayashi,M. and Shinzaki,K.				
TITLE	Transgenic plants carrying neoxanthin cleavage enzyme gene				
JOURNAL	Patent: EP 116794-A 1 18-Jul-2001;				
FEATURES	Riken (Jp)				
source	Location/Qualifiers				
	1..1752				
	/organism="Arabidopsis thaliana"				
	/db_xref="taxon:3702"				
CDS	1..1752				
	/note="unnamed protein product"				
	/codon_start=1				
	/protein_id="CAC41192.1"				
	/db_xref="GI:14347190"				
	/translation="MVSILTMPMSGGIKTPQAOIDIGFPIKRPVYKTOVDVY ELTKROLPTPTTATPPQHRLPLNLFQKAALAIIDAERALISHEDSPLEPTADP RVQITAGNSIPSESSVRNLIVESTIGDIDGVIRKGANMEPTVGHLTFDDGMV HAAVITNGSASVACRFKTERLYOEKRGRVPFKAGELHGSHGARLMIFYARGLC GLINNONGPVANAGLVFYNNRLLAMEDDLPLYOLKTOTGDLTGVRYPDCOLKSA MIARKLDPVKTELHALSYDVKKPKPYLKFEPSDGVSPELEPLETPIIHDFALT ENFVAPIDDOOVFLKEMISGSKSPVDFEKSRLGIMPKDATEASOIIWNSPEFC FILMANESPEPEREIVYGICSMSPADSIFGNRDSLSVSEIRINARTRTTRSLIL VNEDEVNIIEIGVNWNRIKRTREFAFLAIFWPVSGAKVDLTGEKKTIYGEKYK GGEEFFLPNGNSEENEENDGYIECHVAHEDETKTSQEIINAVALKELETIPSRAVP YGFHGTGYSNELVDOL"				
BASE COUNT	484 a	396 c	428 g	444 t	
ORIGIN					
Query Match	41.0%; Score 738.2; DB 6; Length 1752;				
Best Local Similarity	68.2%; Pred. No. 1.5e-208;				
Matches 1063; Conservative	0; Mismatches 478; Indels 18; Gaps 2;				
OY	244 aacactaacgatgaatttgttcagaagagagggcgccgcgcttggaagcggcgag	303			
Db	193 AATCTCTGCCGGCTTAACAATCTTCAGAAGCGGGCGATTGCATCGACCGGCTGAG	252			
OY	304 gttcccttgtcaagcagaagaagtacaaccgcgtcttaaaagcgtatcctagtgt	363			
Db	253 CGTGATTAATCTCACAGGAGCAATACTCTCCACTTTCCAAAACCGCTGATCCAGCTGT	312			
OY	364 cagatcgccggaatatctgtctcgcgtgaatgacacgccgctccgcgcgtaatcttcggtg	423			
Db	313 CAGATTGCCGGGAATTATTTCCCCGGGTACCGGAATCTTCGTCGGCGGAACCTCACCGTC	372			
OY	424 gtcgaaaaactcccgatctcatcaaaagagtgtatgtgycgaacggagctaaccact	483			
Db	373 GAAGGAACAATCCCAGCATGCAATTACAGGTGTTTATATCCGTAAGGGGCGCAATCCGATG	432			
OY	484 cacgcgcgcgttagcaggtcaccaactctcttcgacggaagcgtatggttcacgcgctcaa	543			
Db	433 TTTTGAGCCAACAGCTGCGGACCACCATTTATTTTCAGCGAGACGGAATGTTTCACGACTTAA	492			
OY	544 ttcgaacacggttcagctagctacgctgtgcggtttactcagactaacggtttgttcag	603			
Db	493 ATTAACAACAGGTTACGTACGTACCATGCGCGTTTACAAAAACCGAGAGATTGGTTCAG	552			
OY	604 gaacgtcaattggcgcgacccggtttccccaaagccatcggatggaacttcacggccaacc	663			
Db	553 GAAAAACGATTGGGTGACCAAGTTTTTCCCGAAGAACATCGCGGACCTTCACGCTACTCG	612			

OY	664	ggtatctgcgcgaactaactgatacttcaatgcagagctgcaacgcglatagtcgaaccgcgca	723
Db	613	ggaattccgacgctttgattgctgttttacgcacgctggcgcttctgtgatcaacaccaa	672
OY	724	caacgaacccggttgatagcttaaacgccggttttgctcatttcaatggccggtatgtgcata	783
Db	673	AACGGCGTGGAGATGAGCAAAACGCCGGTTTGATTACTTTAATAACCGCGCTTTTAGCTATG	732
OY	784	tcggagagatgtattacattcaacaggttcagatacctccatccatbtgagatttaaacacgct	843
Db	733	TCAGAGACGATTTTACCGTACCAATTAAATTACTAAACCGCGCATCTCCAAACCGTT	792
OY	844	ggtcgttccgaattttgatbtgacaattagaaatccacaatgtatgccaccgccaaagtcgac	903
Db	793	GGACGTTACGATTTTCAGACGGTCACTTAAATCCGCAATGATAGCTACCCGAACTGGAC	852
OY	904	ccggaatccggttgaaactcttcgctttaaagtaagagtcgtttcaaaagcttaactaa	963
Db	853	CCGGTTACGAAGGAGGACTTCAACGCCGTTAAGCTACACACCTCGTTAAGAAACCTTAACGTGA	912
OY	964	tacttccgattctcaacggaaggaacccaataaccggaacgctgagatctgactgtatcg	102
Db	913	TACTTCAATCTTCGCCACAGAGCGGCTTAAATCGCGGAATTTGGAGATCCCGCTCGAACT	972
OY	1024	ccaacgaatgcacacgaatttcgcgattacagagaactctgctgcgtacctgcacga	108
Db	973	CCGACGATGATTTCAACGATTTGGCTATTAACGGAATAATTTGTGGTGAATCTGTATCAACA	103
OY	1084	gtcgttttcaagctgcccggagatgatccgcggtgggtgtcctcggbtgttatacgaaagac	114
Db	1033	GTCCGTCTCAAGCTCCGCGAGATGATTTCCGGTAAATCTCCGGTGTTTTCGACGAGAA	109
OY	1144	aaggtccgaagatccggaatttgaaacaatacgcgaagaatcatcatgaacattaaing	120
Db	1093	AAGGTTTCCCATTTGGGGATATATCCCAAGACCGACAAAGCTTCTCAGTAAATCTG	115
OY	1204	attgattgtccagattgttctgtcttcacatctctcgaagacgttggagaagccagaa	126
Db	1153	GTGACACTCCCGGAGACGCTTGTTTTCATCTCTCGGAATGCAATGCGAATGCGCGGAGAC	121
OY	1264	gatgaagtcgtcgttgataggttctctgtatgatctccacagaactcaatttcaagagct	132
Db	1213	GAGGAGATTTGGTGATCCGATCCGTATATGTCGCCGCGGATTCATCTTCAACGAGAGA	127
OY	1324	gacggaactctcaagaagtgctctcttgaaatccgcctgatatcaaaacggtgatca	138
Db	1273	GACGAGACTTGAGAAAGGTTTGTTCGAGATCAAGATTAACCTTCAGAACACTTAATAAC	133
OY	1384	actccgcctccgatactctccaagaagatcaacaagtcaacctcgaaagcaggaatggtc	144
Db	1333	ACGCGTCTGTTGTTCTTGTTACAGAGAT-----GTAAATTAGGATTTGGTATGTGTT	138
OY	1444	aacagaacatgtctgcgcccgaataacaanaattcgcttaacttgcttaagccgagcgt	150
Db	1387	AACCGGAACCGGTTTAGGAAGAAAAACCGGTTCCGTTTGGCTATTTGCTTATCTCTTG	144
OY	1504	octaaagctcaagatctgcgtcaaaagttgatactcaactcgtgagagaagttaagaacatct	156
Db	1447	CCAAAAAGTTTCCGTTTCGTTCMAAGCTGCATTTTGACACGGTGAAGTGAATAAATATAT	150
OY	1564	tacgcgcgaataacgcttaccgagagagagcctgtcttctcccg-----agaa	161
Db	1507	TACGGCGGTGAGAAATATAGCGGGCGGAACCGTTTCTTGGCCCGCAACTCCGTTAACGGC	156
OY	1612	ggaagagagaaagacgaagatacatcctctgtttcgttcaacgagaaacatgynaa	167
Db	1567	GAAAGAAATGAAAGATGACAGGTTATATTTTGTTCACCTTCATGACGAAGAAACAAAGACA	162
OY	1672	tcggagttacagatagttaacgcgcgttagctagaggttgaagcaaggtttaaacttcog	173
Db	1627	TCAGAGCTTCAGATTTATTAACGCTGTATTAATTTAAAGCTTGAAGCTACGATTTAAACTACCG	168
OY	1733	tcaaggtctccgaacttcaacgttaccatctcgaatccggaatccggaatgtttggcgaaga	1790

Db	1687	CTAGAGTACGCTATGGCTTTCATGCACATTGTGATTCGATTCGAAAGTACTCGTTGATCA	1745
RESULT	14		
ATP28J12			
LOCUS	110102 bp	DNA	linear
DEFINITION	Arabidopsis thaliana DNA chromosome 4, BAC clone F28J12 (ESSAII project).		
ACCESSION	AL021710		
VERSION	AL021710.1	GI:2832639	
KEYWORDS			
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots: Rosidae: eurosids II: Brassicales; Brassicaceae; Arabidopsis.		
AUTHORS	1 (bases 1 to 110102)		
JOURNAL	Bevan, M., Hilbert, H., Braun, M., Holzer, E., Brandt, A., Duesterhoeft, A., Bancroft, I., Mewes, H.W., Mayer, K. and Schueller, C.		
REFERENCE	2 (bases 1 to 110102)		
AUTHORS	EU Arabidopsis sequencing project.		
JOURNAL	Submitted (03-FEB-1998) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk		
FEATURES			
source	Location/Qualifiers		
gene	1..110102		
exon	/organism="Arabidopsis thaliana"		
	/variety="Columbia"		
	/db_xref="taxon:3702"		
	/chromosome="4"		
	1988..3739		
	/gene="F28J12.10"		
	1988..3739		
	/gene="F28J12.10"		
	1988..3739		
	/number=1		
	1988..3739		
	/gene="F28J12.10"		
	/note="strong similarity to neoxanthin cleavage enzyme, Lycopodium esculentum, PATCHX:E325797; and viviparous-14, Zea mays, PATCHX:G2232017"		
	/codon_start=1		
	/product="neoxanthin cleavage enzyme - like protein"		
	/protein_id="CA16715.1"		
	/db_xref="GI:2832640"		
	/db_xref="SPTREMBL:O49505"		
	/translation="MVSLLTTPMSGIKTPWPAQIDLGFRPIKRPKYIKTVQIDVT ELTKRQLETPTTATPPQHNPLRLNIFOKAAALIDAERALISHEODSPPKTADP RVOLAGNSPVPESSVRNRLTYEGRIIDPGICGVYIRNGANPMEPTAGHLLPDGGMV HAVKTTNSASACREFTKTERIVOKRIGRPYFPAISELHSHSIAKMLFYAGLC GLINNQNGVANAAGLVTFNNKLAMSEDDLEPYQKITQIGDQVTVGRDPEGQKSA MIAHPKLDVTKELHLSYDVYKKPYLKYFRSPDGVKSPELEIPLEPTMIDHFAIT ENFVYIPDOOVVFKLGEIMISGKSPVFDGEEKSRIGMPKDATEASQIIMVSPETFC FHLNMAESPETEELIVIGSCSPADSIJENEDSELSVLSIRINLRKRTKRSIL VNEVDNLEIGWNRNLRGRKTRFALTAIPMPKSGFAKVDICTGEMKKYVGEEXY GGEFFLPNGSNGENEDDGIIFCHVHDEERKTSLEIDINAVNLKLEATIKLPSPRV YGFHGTVDNSMLVDQL"		
	/complement(join(5457..5645,5970..6074,6165..6265,6367..6436,6535..6597,6677..6743,6822..6946,7021..7098,7188..7234,7348..7472,7566..7702))		
	/gene="F28J12.20"		
	/complement(join(5457..5645,5970..6074,6165..6265,6367..6436,6535..6597,6677..6743,6822..6946,7021..7098,7188..7234,7348..7472,7566..7702))		
	/note="strong similarity to (S)-2-hydroxy-acid oxidase (EC 1.1.3.15), peroxisomal, Spinacia oleracea (spinach), PIR1:OXSPH		
	contains EST gb:788637, N37256"		
	/codon_start=1		
	/product="glycolate oxidase - like protein"		
	/protein_id="CA16716.1"		
	/db_xref="GI:2832641"		
	/db_xref="SPTREMBL:O49506"		
	/translation="WEITNMEYERIAEKLEPKVYDIYVYASGAEQDQWLTQENRNAFSR ILFERPILIDVSGIDVSTTVLGFINISMPIMAPTMQKMAHPDGLAARATSAAGTI MTLSSMATICVSEVASTGPGIRFQLYYKIDRNVYIQVRAEAGFKAIALVTPTP R LGRRESIDIKNRPALPRGLTKMFEGLDGLKIDKTDNSGLASVYAGVOQSLSMKD IKM LOSITSLPLVKGVITARDARVAVEGAAGITKDSNMGAROLDYVPATVIALEEVKAV EGRITPFLDGGVRRGTDYFKALALGASVFPVGRPSLPSLADGEGVAKMQLMDEP ELTMALSCRSRLREISRTHITKTDWDPHTLSAKL"		
	/complement(5460..5645)		
	/number=1		
	5460..7702		
	/gene="F28J12.20"		
	/complement(5646..5969)		
	/gene="F28J12.20"		
	/number=1		
	/complement(5970..6074)		
	/gene="F28J12.20"		
	/number=2		
	/complement(6075..6164)		
	/gene="F28J12.20"		
	/number=2		
	/complement(6165..6265)		
	/gene="F28J12.20"		
	/number=3		
	/complement(6266..6366)		
	/gene="F28J12.20"		
	/number=3		
	/complement(6367..6436)		
	/gene="F28J12.20"		
	/number=4		
	/complement(6437..6534)		
	/gene="F28J12.20"		
	/number=4		
	/complement(6535..6597)		
	/gene="F28J12.20"		
	/number=5		
	/complement(6598..6676)		
	/gene="F28J12.20"		
	/number=5		
	/complement(6677..6743)		
	/gene="F28J12.20"		
	/number=6		
	/complement(6744..6821)		
	/gene="F28J12.20"		
	/number=6		
	/complement(6822..6946)		
	/gene="F28J12.20"		
	/number=7		
	/complement(6947..7020)		
	/gene="F28J12.20"		
	/number=7		
	/complement(7021..7098)		
	/gene="F28J12.20"		
	/number=8		
	/complement(7099..7187)		
	/gene="F28J12.20"		
	/number=8		
	/complement(7188..7234)		
	/gene="F28J12.20"		
	/number=9		
	/complement(7235..7347)		
	/gene="F28J12.20"		
	/number=9		
	/complement(7348..7472)		
	/gene="F28J12.20"		
	/number=10		
	/complement(7473..7565)		
	/gene="F28J12.20"		

Query Match	Best Local Similarity	Matches 1063	Conservative	Score 738.2	DB 8	Length 110102
41.0%	68.2%	0	Mismatches 478	Indels 18	Gaps 2	
244	aacactaacagatgaattggtccagaagacgagcgagcagcglttgacgacgagcgag	303				
Db	2180	AATCCTCTCCGGCTAAACATCTTCCAGAAAGGCGGCGAATTCGATCGACGCGCTGAG	2239			
304	ggttcctgtgtcagcagcagagagcgtacacccgcgtctcctaagacggtcgtatgctt	363				
Db	2240	CGTGCATTAACTCCACAGAGCAAGATTTCTCCACTTCCCAAAAACGCGATCCAGGTGT	2299			

QY	364	cagatcgccggaaatttctgtccggtggaatgaacagcccgctccggcgtaattcttcggtg	423
Db	2300	CAGATTGCCGGGAATTATTTCCCGCGATACCGGAATCTTCGCTCCGGGAAACCTCACCGTC	2359
QY	424	gtcgaaacttcccgattccatccaagaaggtatgtgcgaacggagctaaaccactt	483
Db	2360	GAAAGAACATTCCTGACTGCTATTTGACGGGTGTTTATTCCTGTACGGCGCAATCCGATG	2419
QY	484	cacgacccggtgaacaagltcaaccacttcttcgaagagacggtatgttcaagccgtcaaa	543
Db	2420	TTTGAGCCAAACAGCTGGGACACATTTATTTGACGAGACGGAATGGTTACGCAGTTAA	2479
QY	544	ttcgaaacccggtltaagctaaagctgcgttcgggttctaactcaactaaecggittgttaag	603
Db	2480	ATAACCAACGGTTTACGTACTACGCTACGCTGCGGTTTACAAAAACCGAGATTGGTTACG	2539
QY	604	gaacgtaattgggttcgaacccggtttccccaagaacatacgtggaacttcaagcccaacc	663
Db	2540	GAAAAACGATTGGGTTCACACAGTTTTCGAAAAGCAATGGGAGGCTTTCACGCTACATCG	2599
QY	664	ggtatgtcccgacccaacgcatctatcaacgcagaagctgcagccggtatagtcgaccgga	723
Db	2600	GGAAATCGACCTTTGATGTCGTGTTTAAACGACGATGGGCTTTGATGCTGATCAACAACCA	2659
QY	724	cacggaacccggtltagtcaaaccccggtttgtctatlttaatgagccggtltagtcatg	783
Db	2660	AACGCGCTCGGAGTAGAACAAACGCCGGTTTGTTTACTTTAAATACCGCGCTTTTACCTATG	2719
QY	784	tcgagagttgatttaactcttccaaagttccagatcaaccccccaatggagatttaaaacggt	843
Db	2720	TCAGAAAGCGATTTACCGCTTACCAATTTAAATATTCGAAAACCGGCGATTCACAAACGTT	2779
QY	844	ggtcgtgttcgaatttttgaatggaacaatgaatcaacaaatgatgtgccaccgaagaatgcac	903
Db	2780	GGAGCTTACGATTTTCGACGGGTACAGTTAAATCCGCAATCATATGCTCACCCGAACCTGGAC	2839
QY	904	ccggaatccggttgaactcttcgcgtttaaagctacagcgttcgttccaagccttaactaaa	963
Db	2840	CCGGTTTACGAAGGAGCTTTCACGGGTTTAACTACGAGTGTGTTTAAAGAAACCTTACCTCGTAA	2899
QY	964	tacttcgattctcaaccggagcgaactaaataccggaacgctgagatcaagtttaactga	1023
Db	2900	TACTTCACATTTCTGCCACAGACGGGTTTAAATCCCGGAATTTGGAGATTCCTCGTCAAACT	2959
QY	1024	ccaagaagatgcacagatlttcgcgattacagaaacttcgctgcgtacatccagaacaa	1083
Db	2960	CCGAGACGATTCACGATTTCCGTATACGGAATAATTTGTGTGATTTCCGTATCAACAA	3019
QY	1084	gtcgttlttcaagctgcgcggagatgatccgcggttgcgttcggttgcgttgcgttgcgttgc	1143
Db	3020	GTCGCTTCAAGCTCGCGCGAGATGATTTCCGGTAAATCTCCGGTTGTTTTCGACGAGAA	3079
QY	1144	aaggttcgaagaatcccggtattttagaataataacgcggaagatctatcgaatactaaatg	1203
Db	3080	AAGGTTTCCCATTTGGGGATTAATGCCCAAGACGCAAGAAAGCTTCTAGTAAATCTGG	3139
QY	1204	attgatgtccagaattgtctctgccttccatctcttgaaacgcttgggaagagccagaanaa	1263
Db	3140	GTGAACCTCTCCGGAGACGTTCTGTTTTCATCTCTGGAATGTCATGGGAATCCGCGGAGAC	3199
QY	1264	gatgaagtcgctgcgtatagaaggtccctgtatgatctcaaccagactcaatttccaagactc	1323
Db	3200	GAGGAGATTGTHGATTCGATCGATCGATGCTGATTCGCCGGCGGATTCATTTTCACAGAGA	3259
QY	1324	gaacgaagaatctcaagaagtgtcctgcttgcgttgaataccgcctgaatctcaaaacccggt	1383
Db	3260	GACGAGACTTTGAACAAACGTTTGTTCGAATAGAGATTAACCTCGAACAACGATTAACCC	3319
QY	1384	actgcgcgtcgaatcatctcccaagaagatcaacaagtccaactcgaacgagcaggtgtgtc	1443
Db	3320	ACGCGCTTCCTGTTGTTGTTTAAACGAGAT-----GTAAATTTAGACATTTGCTATGGTT	3373
QY	1444	aacgaagaacatgcctcgcgcgtlaaaccnaattgcgtacttgcgtttagccgaagccgtgc	1503

Db 3374 AACCGAACGGTTAGGAAGAAAACCCGGTTCGGTTTGGCTATTCCTTATCCTTGG 3433
1504 cctaagatctcagatcgaagtcgaagtcactactcaggaaggaagtaagaacatctt 1563
3434 CCAAAAGTTTCGGGTTTGGCTAAGSTCGATCTTTGACACCGGAGATGAAAAATATATT 3493
1564 tacgagcgaatacgcgttaaggaagagcctctgttctcccg------agaa 1611
3494 TACGCGGTGAAATATAGCGCGACCGGTTTCTTGCCCGCAACTCCGTAACGCG 3553
1612 ggaagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 1671
3554 GAAAGAAATGAAGATGACGGTTATATTGTCACGTTCAAGCAAGAAACAAAGACA 3613
1672 tcggaggtaccagatagatgaagtcagtagaggtgaaggaaggaaggaaggaag 1731
3614 TCAGAGCTTCAGATTATTAACCTGTATTAAGCTTTGAAGCTTAACGATTAACTACCG 3673
1732 tcaagggtccgtaacgattcagcgtatcatcgcagcagcagcagcagcagcagcag 1790
3674 TCTAGAGTACCGTATGCTTTCATGTCACATTTGTCGATTCGAATGAACACTCCTGATCA 3732

RESULT 15

LOCUS ATCHRIV48 194143 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 48.
ACCESSION AL161548
VERSION AL161548.2 GI:7268604
KEYWORDS
SOURCE
ORGANISM

thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
AUTHORS 1 (bases 107966 to 194143)
Hilbert, H., Braun, M., Holzer, E., Brandt, A., Duesterhoeft, A.,
Mewes, H.W., Lemcke, K. and Mayer, K.F.X.

JOURNAL
AUTHORS 2 (bases 1 to 38835; 28388 to 111084)
Murphy, G., Ridley, P., Hudson, S., Mewes, H.W., Lemcke, K. and
Mayer, K.F.X.
Unpublished

JOURNAL
AUTHORS 3 (bases 1 to 194143)
EU Arabidopsis sequencing, project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@brc.ac.uk

COMMENT
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
this fragment has an overlap with ATCHRIV47 at the 5' end and an
overlap with ATCHRIV49 at the 3' end.

FEATURES
source
Location/Qualifiers

1..194143
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
6377..9293
/gene="AT4g18120"
join(6377..6796,7122..7238,7347..7501,7578..8456,
8543..8627,8716..8784,8871..9293)
/gene="AT4g18120"
/note="Similarity to Me12-like protein, Arabidopsis
thaliana, gb:D86122
contains EST gb:W43784, A1998063.1"
/codon_start=1

/product="putative protein"
/protein_id="CAB78814.1"
/db_xref="GI:7268605"
/translation="MASGDHFGSGIGNMLDGDSPHIGNMLPDDEEELFSGLDLGN
LSSLPARTLDDLEDYDLGSGGLFLEDDPYSLKRGSRMGPASNDNVNRPQITDN
GVGSIACGHPYGEHPISRLTLYRNTNSNVESELDALFEDNSENKDVNGTILVYNLAP
SVSNRDLNLEIGVYGEIKREQLRETPNKRHHKFEVEFDVRSADALAKLNRLEIGKRI
KLEHSRPGGARNMMLQNPLEDODSYSLNHVESPLASSPIGNWRNSP IDHPLOSF
SKSP IFGNLSPTKNI RYPEFSMKTASVNDDEGRFESHLDL FSSSYNNASHKASTF
OOPSGFSGVSSGLNSHPSEHVTLSGSEFLMGSPSSAMVNPFSNREHREPYS
ONGSIHOLHHIGSAPSGCFPRSPETSSMGSAFPGAGGNMAORNLRETSNPKMIS
APRSQLEFTNGSYLMPRATVYSTIDPLEDSNQPDSNGNDIKIOPOLDKIMR
GEDPRTLMITKINIPNKYTRNMLLAIDKNSGTIDFLYLPIDEFNKCNVGAFTNWS
PKFTIAYEAFNGKRMKPFNSEKVASLARIQKALIAHFGNSSLINEDRRCOPIT
FDGSESKYPVNNLLQCTHDLKMIYLFVYINMNLTHCLFETRSSGEFARASNTG
FLIFLSARTHIKADSOVHVHVASGLKKRYQLG"
6377..6796
/gene="AT4g18120"
/number=1
6797..7121
/gene="AT4g18120"
/number=1
7122..7238
/gene="AT4g18120"
/number=2
7239..7346
/gene="AT4g18120"
/number=2
7347..7501
/gene="AT4g18120"
/number=3
7502..7577
/gene="AT4g18120"
/number=3
7578..8456
/gene="AT4g18120"
/number=4
8457..8542
/gene="AT4g18120"
/number=4
8543..8627
/gene="AT4g18120"
/number=5
8628..8715
/gene="AT4g18120"
/number=5
8716..8784
/gene="AT4g18120"
/number=6
8785..8870
/gene="AT4g18120"
/number=6
8871..9293
/gene="AT4g18120"
/number=7
complement(9594..9837)
/gene="AT4g18130"
/number=1
complement(join(9594..9837,9944..10237,10323..11130,
11238..13230))
/gene="AT4g18130"
9594..13230
/gene="AT4g18130"
complement(join(9594..9837,9944..10237,10323..11130,
11238..13230))
/gene="AT4g18130"
/note="Contains ATP/GTP-binding site motif A (P-loop)
AA634-641:Phytochrome chromophore attachment site
AA317-326
contains EST gb:R21020, AA395462"
/codon_start=1
/product="phytochrome E"
/protein_id="CAB78815.1"

[illegible][illegible]

```
||||| | |||| ||||| | ||| ||| ||| | ||| |||
Db 111045 AAGGTTTCCCGATTGGGATTAATGCCCCAAGACGACAGAAAGCTTCTCAGATTAATCTGG 111104
QY 1204 attgatgtcccaagattgcttcgtccatctctgaaagcgttggaaagccagaaca 1263
Db 111105 GTGAACCTCTCCGAGACGTTCTGTTTCATCTCTGGAATGCATGGGAATCGCCGGAGACG 111164
QY 1264 gatgaagtcgtcgtataagtgctcgtatgactccaccagactcaattccaagagctc 1323
Db 111165 GAGGAGATTGTGGTGTATGCGATGCTATGTCGCCGCCGATTTCATCTTCAACGAGAGA 111224
QY 1324 gacgagaatctcaagagtgctcctgtcgaatccgcctgaatctcaaaacggttgaatca 1383
Db 111225 GACGAGAGCTTGAAGAGCTTTTGTGCGAGATCAGGATTAACCTCAGAACACGTTAAACCC 111284
QY 1384 actgcgcgtccgactcccaagagaagacaagaagtaacctggaagcaggatgctc 1443
Db 111285 ACGCGTCGTTGCTGTTGCTTAACGAGAT-----GTAAATTTAGAGATTGGTATGTT 111338
QY 1444 aacagaaacatgctcggccgtaaaaccaaatcgccttacttgctttagcgcgagccgttg 1503
Db 111339 AACCGGAACCGGTTAGGAGAGAAACCGGTTTCGGCTTTTGCTATTCCTTATCCTTGG 111398
QY 1504 cctaaagtcctcaggatcgcctaaagltgaltcactactcgaagaagttaagaacaatctt 1563
Db 111399 CCAAAAAGTTTCCGCTTCCCTAAGGTGATCTTTGCACCGGTGAGATGAAAAAATATATT 111458
QY 1564 tacggcgataaccggttaaggagagagcctcgttctcctcccg-----agaa 1611
Db 111459 TACGGCGGTGAGAAATATGCGCGAACCGCTTTTCTTCCCGGCACACTCCGGTAACGCC 111518
QY 1612 ggaaggagagaaagacgaagatacaltcctcgttctcgaagcgaagaagacatgaa 1671
Db 111519 GAAGAAATATGAGATGACGTTATATTTGTCACGTTTCATGACGAGAAACAAAGACA 111578
QY 1672 tcggaggttcaagatagttaaagccgttagcctagaggttgaagcaaggtttaacttcg 1731
Db 111579 TCAGAGCTTCAAGATTATTAACGCTGTTAATTTAAAGCTTAAGCTTAACATPACCG 111638
QY 1732 tcaaggttcgtaaggttcaaggtatcatcatcgcgaagcgaatttggcgaagca 1790
Db 111639 TCTAGAGTACCGTATGGGTTTCAATGACACTTTTGATTCGATTAATCACTCGTTGATCA 111697
```

Search completed: July 24, 2002, 07:42:25
Job time: 7835 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 04:30:44 ; Search time 224.85 Seconds
(without alignments)
13744.479 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800
Sequence: 1 atgctctcttcacgcgcacac.....tggcgaagcagtcgtgtga 1800

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_032802:*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1800	100.00	1800	22	AAD09396
2	824.6	45.8	1818	22	AAD09401
3	798.4	44.4	1839	22	AAD09399
4	738.2	41.0	1752	22	AAD09394
5	606.6	33.7	1815	22	AAD09400
6	528.2	25.3	1734	22	AAD09398
7	229.8	12.8	443	21	AAC565678
8	204	11.3	492	22	AAC82706
9	188	10.4	372	21	AAC565695

10	171.8	9.5	325	21	AAC56548	Eucalyptus grandis
11	169.6	9.4	1788	22	AAC42989	Arabidopsis thaliana
12	169.6	9.4	1788	22	AAD09395	Arabidopsis thaliana
13	169.2	9.4	1950	22	AAE77206	CDNA encoding sunf
14	149.4	8.3	491	21	AAC57157	Pinus radiata tran
15	143.4	8.0	386	21	AAC57162	Pinus radiata tran
16	140.4	7.8	326	21	AAC57167	Pinus radiata tran
17	139.8	7.8	1617	22	AAD09397	Arabidopsis thaliana
18	139.8	7.8	1617	22	AAD09402	Arabidopsis thaliana
19	139.8	7.8	1777	22	AAC36083	Arabidopsis thaliana
20	137.2	7.6	329	21	AAC57165	Pinus radiata tran
21	121.6	6.8	398	21	AAC57145	Pinus radiata tran
22	118	6.6	340	21	AAC56520	Eucalyptus grandis
23	113	6.3	238	21	AAC57169	Pinus radiata tran
24	109.6	6.1	246	21	AAC57168	Pinus radiata tran
25	106.6	5.9	412	21	AAC57164	Pinus radiata tran
26	89	4.9	190	21	AAC57163	Pinus radiata tran
27	78.6	4.4	303	21	AAC57125	Pinus radiata tran
28	76.8	4.3	308	21	AAC57142	Pinus radiata tran
29	74.8	4.2	395	21	AAC57117	Pinus radiata tran
30	67.8	3.8	285	21	AAC56908	Pinus radiata tran
31	65.8	3.7	567	21	AAC56546	Eucalyptus grandis
32	64.6	3.6	457	21	AAC56910	Pinus radiata tran
33	63	3.5	445	21	AAC56948	Pinus radiata tran
34	62.4	3.5	436	21	AAC57124	Pinus radiata tran
35	58.2	3.2	304	21	AAC56973	Pinus radiata tran
36	56.6	3.1	419	21	AAC56527	Eucalyptus grandis
37	56.6	3.1	447	21	AAC56480	Eucalyptus grandis
38	56.6	3.1	508	21	AAC56511	Eucalyptus grandis
39	55.2	3.1	384	21	AAC56540	Pinus radiata tran
40	53.8	3.0	311	21	AAC57151	Pinus radiata tran
41	49.4	2.7	321	21	AAC56510	Eucalyptus grandis
42	48.4	2.7	450	21	AAC56474	Eucalyptus grandis
43	46.4	2.6	501	21	AAC43547	Zea mays DNA fragm
44	45	2.5	384	21	AAC56556	Eucalyptus grandis
45	44	2.4	311	21	AAC56536	Eucalyptus grandis

ALIGNMENTS

RESULT 1	
ID	AAD09396 standard; cDNA; 1800 BP.
AC	AAD09396;
XX	
DT	10-SEP-2001 (first entry)
XX	
DE	Arabidopsis thaliana neoxanthin cleavage enzyme, ATNCE3 CDNA.
XX	
KW	Neoxanthin cleavage enzyme; ATNCE3; abscisic acid; ABA; herbicide;
KW	stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW	plant growth protectant; ss.
XX	
OS	Arabidopsis thaliana.
XX	
EH	Key
FT	Location/Qualifiers
FT	1..1800
FT	/tag= a
FT	/product= "Arabidopsis thaliana ATNCE3 protein"
XX	
PN	EP116794-A2.
XX	
PD	18-JUL-2001.
XX	
PF	11-JAN-2001; 2001EP-0300218.
XX	
PR	13-JAN-2000; 2000JP-0010056.
PR	11-JAN-2001; 2001JP-0003476.
XX	
PA	(RIKE) RIKEN KK.
XX	

PI Iuchi S, Kobayashi M, Shinozaki K;
XX WPI: 2001-400081/43.
DR P-PSDB; AAE04784.
XX
PT A DNA encoding a protein with a neoxanthin cleavage activity for
producing transgenic plants with improved or decreased stress tolerance
PS
XX
PS Claim 3; Page 32-36; 101pp; English.

CC The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present cDNA sequence encodes
CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED3 protein.
CC The AtNCED3 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC library using a cDNA of the CPRD65 (Copeia Responsive to Dehydration)
CC gene isolated from copeia plant as a probe.

XX
XX Sequence 1800 BP; 458 A; 464 C; 439 G; 439 T; 0 other;

Query Match 100.0%; Score 1800; DB 22; Length 1800;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtctcttcacggaacgctgcgttctctggaagatgcttggtggcaatcact 60
DB 1 atggtctcttcacggaacgctgcgttctctggaagatgcttggtggcaatcact 60
QY 61 cagcgccattatcgtcttctcgaagctccgacttgatgtatgtcttactatg 120
DB 61 cagcgccattatcgtcttctcgaagctccgacttgatgtatgtcttactatg 120
QY 121 gcaagtcgttcacagcgaagcgaatgttcacatctcgcctccacagctct 180
DB 121 gcaagtcgttcacagcgaagcgaatgttcacatctcgcctccacagctct 180
QY 121 gcaagtcgttcacagcgaagcgaatgttcacatctcgcctccacagctct 180
DB 121 gcaagtcgttcacagcgaagcgaatgttcacatctcgcctccacagctct 180
QY 181 cattccctaagaatcatcaactcccgcatgtgtttaagcccaagccaaagaa 240
DB 181 cattccctaagaatcatcaactcccgcatgtgtttaagcccaagccaaagaa 240
QY 241 tcaacaactaagaatgattgttccagaagcgcgcgcgaggttgagcgcgcg 300
DB 241 tcaacaactaagaatgattgttccagaagcgcgcgcgaggttgagcgcgcg 300
QY 301 gaaggttctcttcacagcgaagcgaatcaccgcgttcttaaaagcggtacgt 360
DB 301 gaaggttctcttcacagcgaagcgaatcaccgcgttcttaaaagcggtacgt 360
QY 361 gtccaagtcgcggaatttgcctcgggtgaatgaacagccgctccggttaactc 420
DB 361 gtccaagtcgcggaatttgcctcgggtgaatgaacagccgctccggttaactc 420
QY 421 gtgttcggaagaacttccggttcacataaagggtgatgtgccaagcgaagtaacca 480
DB 421 gtgttcggaagaacttccggttcacataaagggtgatgtgccaagcgaagtaacca 480
QY 481 ctccaagcgcgttgacagatcaccactcttcgaacggaacggtatggttcacgcgtc 540
DB 481 ctccaagcgcgttgacagatcaccactcttcgaacggaacggtatggttcacgcgtc 540
QY 541 aaatcgacaacggttcagctagctacgcttgcggtttactcagactaacgcggtt 600

DB 541 aaatcgacaacggttcagctagctacgcttgcggtttactcagactaacgcggtt 600
QY 601 caggaacggtcaattggttcgaccggttttccccaagaacatcggtatgttaacgcgcac 660
DB 601 caggaacggtcaattggttcgaccggttttccccaagaacatcggtatgttaacgcgcac 660
QY 661 accggtatgcccgaactcagctatctcagcagagctcgaacggtatagtcgaccgcg 720
DB 661 accggtatgcccgaactcagctatctcagcagagctcgaacggtatagtcgaccgcg 720
QY 721 gcaacggaacggtgtagctaaagcgcggttggctatttcaatggcggtattggt 780
DB 721 gcaacggaacggtgtagctaaagcgcggttggctatttcaatggcggtattggt 780
QY 781 atctcggaagatatttactctcaccaggtcacaatcaccatccaatggatttaaaac 840
DB 781 atctcggaagatatttactctcaccaggtcacaatcaccatccaatggatttaaaac 840
QY 841 gtgttcggttcgatttgaatgacaatgaatccacaatgattgccaccggaagtc 900
DB 841 gtgttcggttcgatttgaatgacaatgaatccacaatgattgccaccggaagtc 900
QY 901 gaccggaatcgcgtgaactcttcgtttaagctaacgagctggttcaaaagccttacct 960
DB 901 gaccggaatcgcgtgaactcttcgtttaagctaacgagctggttcaaaagccttacct 960
QY 961 aaatcttcggtatcttcacagcgaagcgaatcaatccagcgtgagatgttcgctgt 1020
DB 961 aaatcttcggtatcttcacagcgaagcgaatcaatccagcgtgagatgttcgctgt 1020
QY 1021 cagccaacgataatgacgaatttcgcatcagaagaacttcgtctgcttacctgacag 1080
DB 1021 cagccaacgataatgacgaatttcgcatcagaagaacttcgtctgcttacctgacag 1080
QY 1081 caagtcgttctcgaagctgcgcgagatgatacgcggtggtcttcggtttacgacaag 1140
DB 1081 caagtcgttctcgaagctgcgcgagatgatacgcggtggtcttcggtttacgacaag 1140
QY 1141 aacaagtcgcaagatctgggattttagaacaataacgcgaagaattcatcgaacatt 1200
DB 1141 aacaagtcgcaagatctgggattttagaacaataacgcgaagaattcatcgaacatt 1200
QY 1201 tggattgatgtcacaagattctctcgtcctccacatctctggaacgcttgggaagccagaa 1260
DB 1201 tggattgatgtcacaagattctctcgtcctccacatctctggaacgcttgggaagccagaa 1260
QY 1261 acagatgaagtcgttcgtatagaggtccctgtatgaactccacgaactcaatttcaacag 1320
DB 1261 acagatgaagtcgttcgtatagaggtccctgtatgaactccacgaactcaatttcaacag 1320
QY 1321 tctgcagaagatctcagaatgtccgtctggaatccgccttgatctcaaaacgcgtgaa 1380
DB 1321 tctgcagaagatctcagaatgtccgtctggaatccgccttgatctcaaaacgcgtgaa 1380
QY 1381 tcaactcgcgttcgcatactctcacaacgaagaatcacaacgtcgaagcgaaggtatg 1440
DB 1381 tcaactcgcgttcgcatactctcacaacgaagaatcacaacgtcgaagcgaaggtatg 1440
QY 1441 gtcaacaagaacaatgctcggcgtaaaaccaaatctcgcttacttggcttaagccgag 1500
DB 1441 gtcaacaagaacaatgctcggcgtaaaaccaaatctcgcttacttggcttaagccgag 1500
QY 1501 tggcctaaagtctcaagatctcgttaagtgtatctcaacttggagaaagttaagaaacat 1560
DB 1501 tggcctaaagtctcaagatctcgttaagtgtatctcaacttggagaaagttaagaaacat 1560
QY 1561 cttaacgcgataacggttaacgagagagcctctgtttcccccgggaagagagagag 1620
DB 1561 cttaacgcgataacggttaacgagagagcctctgtttcccccgggaagagagagag 1620
QY 1621 gaagcgaagatacatcctctgttctcgtttcacgacgagaagacatgtgaaatcagaatta 1680

Dd	1621 gaagacgaagatacaatctccttgcgttcacgacgagaagaaacatgtaaatcgaggacta	1680
Oy	1681 cagaatgaacgcgttagcttagcttagtgattgaagacaacggtttaactccgtlcaagggtt	1740
Dd	1681 cagaatgaatcaacgcgttagcttagcttagtgattgaagacaacggtttaactccgtlcaagggtt	1740
Oy	1741 ccgaacggatttacggatcatcattccgagaccgatgattggcgaacgaagtctgtgtga	1800
Dd	1741 ccgaacggatttacgaagtatcatcattccgagaccgatgattggcgaacgaagtctgtgtga	1800
RESULT 2		
ID	AAD09401 standard; cDNA: 1818 BP.	
AC	AAD09401;	
DT	10-SEP-2001 (first entry)	
DE	Lycopersicon esculentum neoxanthin cleavage enzyme, LENCED1 cDNA.	
KW	Tomato; neoxanthin cleavage enzyme; LENCED1; abscisic acid; ABA;	
KM	stress tolerance; transgenic plant; plant breeding; antisense-therapy;	
KW	plant growth protectant; herbicide; ss.	
OS	Lycopersicon esculentum.	
FH	Key	Location/Qualifiers
FT	CDS	1..1818
FT	/tag=	a
FT	/product=	"Lycopersicon esculentum LENCED1 protein"
PN	EP1116794-A2.	
XX	18-JUL-2001.	
XX	11-JAN-2001; 2001EP-0300218.	
PR	13-JAN-2000; 2000JP-0010056.	
PR	11-JUN-2001; 2001JP-0003476.	
XX	(RIKE) RIKEN KK.	
PI	Inchl S, Kobayashi M, Shinozaki K;	
PI	WPI: 2001-400081/43.	
DR	P-PSDB: AAEO4789.	
CC	A DNA encoding a protein with a neoxanthin cleavage activity for	
CC	producing transgenic plants with improved or decreased stress tolerance	
CC	-	
CS	Claim 3; Page 67-71; 101pp; English.	
XX	The invention relates to neoxanthin cleavage enzymes and their	
CC	corresponding CDNA molecules. Acid (ABA) biosynthesis under drought stress-	
CC	role in endogenous abscisic acid (ABA) biosynthesis under drought stress-	
CC	Neoxanthin cleavage enzyme is used for improving stress tolerance in a	
CC	plant when expressed in a plant cell. The invention also relates to	
CC	methods for increasing or decreasing stress tolerance in a plant by	
CC	introducing the DNA into the plant, and a transgenic plant into which a	
CC	neoxanthin cleavage enzyme is introduced. The improvement of stress	
CC	tolerance in plants is useful, for example in plant breeding. Neoxanthin	
CC	cleavage enzyme genes are useful for producing transgenic plants. An arid	
CC	land can be improved by growing transformant weed for several years and	
CC	then removing the weed by specifically lowering stress tolerance in the	
CC	weed by inducing an inducible promoter. The present CDNA sequence encodes	
CC	lycopersicon esculentum neoxanthin cleavage enzyme, LENCED1 protein	
CC	related to the invention.	
XX	Sequence 1818 BP; 569 A; 383 C; 369 G; 497 T; 0 other;	

Query Match	45.8%	Score 824.6	DB 22	Length 1818	
Best Local Similarity	69.6%	Pred. No. 2.1e-256			
Matches 1134	Conservative 0	Mismatches 489	Indels 6	Gaps 1	
QY	ctccgcgctctattcttcctctaagcgcatacgcatactctccgcgcatlgtgtttaagccca	229			
DB	115 cttcaaatatataaacacccaagaataabacaatttccacacccaacaaagaacaaaca	244			
QY	230 aagcgaagaatcccaacacataacagatgatattgttccagagagcgggcgagcgt	289			
DB	245 actctctctcttcttccaacttcccaagtggaatttgtgtgaagaagcagagcaatggtc	304			
QY	290 tggacgcggcgagaggttctcctgttcagccacagagaagctaacccggtctctaanaag	349			
DB	305 tagatgctgttagaagaagtgttctaactaaacatgaacttgaacaccccttgcggaacag	364			
QY	350 ctgacatccatgtttcaagatcgcgcggaatttgcctccggttgaacagcccgctccgcg	409			
DB	365 ccgaccaccaagatccagatttctgtggaatttgcctccggttacgggaaataatccagttctc	424			
QY	410 graatcttcggtgtgtgcggaacattcccgatcttccaaatgaagatgtgtatgtgcgaacg	469			
DB	425 aatctcttcgcgtacacgggaaaaataaccacaattgttcaagtcgtttacgtttcgaacg	484			
QY	470 gaagctaaccacatctcacgagccggtgtgacagcgttccacacttcttcgacggagacgtatg	529			
DB	485 gaggctaacctcttlttgacacacacgcgcgacacacattctctgcagcgcgagcgtatg	544			
QY	530 ttcaagccggtcaaatctgcagacacggttcaagctagctagctgtccggtttactcaagcta	589			
DB	545 ttcaagccggtcaaatctcaaaaatgtgtcgcgcagttacgctgtccggtttcacatgtgaacg	604			
QY	590 accggtttgttcaggaaacgttcaattggcgtgcgcacggttttccccaagaagcaltgcgtgac	649			
DB	605 aagagcgtgttcaagaaaaagcttctgggttcgcgccggttttccctaagaagcaltgttgaat	664			
QY	650 ttcaagcgccacacccggtatgtccgcgactatgtctattctacgcgacagctgcagccgcta	709			
DB	665 taactgtacactcttggaatttgcaagcgtatgctgttttaagctcgttgggcctctcgac	724			
QY	710 tagtcgaaccgcgcacacaggaaccggtgtatagctaaagccggttgtgtctattccaatgccc	769			
DB	725 tgtgtgatctacagtaaaagaaactgtgtgtgcacaacgcggtttatgtctattccaataacc	784			
QY	770 gtttttggtctatgttcggggagatgttttaaccttaccaaatltcaagatcaatccaatggag	829			
DB	785 gattctctgctatgttcctgcagagatgatttgccttaccatgttaaaagtaaacccaaccgcg	844			
QY	830 atttataaaacgcttgtcgttgcgttccgatttttgatgtgacaatttgaatccaatgatgtccc	889			
DB	845 atcttataaacagaggttcgatttcgatttcgcagcgcaggtataaaccacatgatatgctc	904			
QY	890 acccggaagtcgaaccgcgaatcccggttgaactcttcggtttaagctacgacgttcggttcaa	949			
DB	905 accccaagctgcacccaagcttcccggttgaagctatttgccttgcctacgatagtgatccaga	964			
QY	950 agaccttaacataatctatcttcggtatcttcacccgagacccaatcaatccgcgcgtcagaga	1009			
DB	965 agccataaccctacagtacttccagatttccaataaaggggaaaaatcaaatgattgttgaaa	1024			
QY	1010 ttcaagcttgatatacgaacgaagatgatgtacagatattccgcgatactacagagaacttcgtc	1069			
DB	1025 ttccagttgaaagacccaacaatgatgtcatgtatttcgcaattactcagaaacttcgctcga	1084			
QY	1070 taactgcacagcaagtcttttccaagctgcgcgagagatgtatccgcggttgggtctccggtg	1129			
DB	1085 ttccctgatcaaaaagctcgttttccaagatgtctcgtgaatgtatccgttggaggtttcacgcg	1144			
QY	1130 tttaagacaagaagacaggtcgcgaagatcggagattttaaacaataacgcgcgaagatcat	1189			
DB	1145 tttaagacaagaagaaagttcccgatcttggtgatctccgataagtaagtcggaagaaggtct	1204			
QY	1190 cgaacattaaagagattgtatgttccagatgttcttgcgttccatctctgaaacgcttgg	1249			

Db	1205	ctgattggaatggtgtggaagttacccgattggtttccgtttcccaaccctcggaaatgcttggg	1266
OY	1250	aagagccgaacacagatgtaagtcgtctgtgaataggcttcctgtatgatactccacacagatctaa	1309
Db	1265	aagaagacgaanaacagatgtaaatcgttgttaatttggtttcatgtatgataccacacagatctca	1324
OY	1310	tttcaacacgagtctgcacgagaaatctcaagaagtgctctgtcttgaaatccgctgtaattcca	1365
Db	1325	ttttcaatgtaattgtgatgtaagggtctaaagaagtgcttttaaccgaaatccgtctccatcttga	1384
OY	1370	aaacgggtgtaatactacccgcgtccgtgatactctccaaagaaatacaagaagccaaccccg	1429
Db	1385	aaacggggaaatcaacaagaanaatccataatcgnaaaacccggatgtaacaagtgaatttag	1444
OY	1430	aagcagagatgtgtcaacaagaacaatgctcgtcgccgttaaaaaccaaattcgttacttggtct	1489
Db	1445	aagctggaatgtgtgaatccgaaacaacaactcggaaagaaacagaatagtgtatttggtcta	1500
OY	1480	tagccgagccgtgtgcctaaagtctcaagattcgtctaaagttgactactactatgagaag	1549
Db	1505	tcgctgtaaccatctgccaagaagttctctgttttgcaaaagttaaacctgttccaccggttgaag	1564
OY	1550	ttagaacaatctttaacgggtataaacctgttaacgagaagagcctctggtttcccccggag	1609
Db	1565	ttgagaatcttcaattatctggtgacacaacaataatgtgtggaaacctcttttttaaccaagag	1624
OY	1610	-----aaggaggagaggaagacgaagaacatactcctcgtttcgttcaacgacgagaaga	1663
Db	1625	accccaacaagaagaagaagaacagatggttatatttttaagcttcttcgttcaacgagtgaag	1688
OY	1664	catggaatctcgagattacagatagttaacgcgttagcttagaaggttgaagacaacggtta	1723
Db	1685	aatggaatctcgaactcgaaattgtttaacgcaatgagttggaagttgagggcaactgtga	1744
OY	1724	aacttcgctcaagggttcgctgaacgatttcaacgttatcaatcaatccggagccgagtatttg	1783
Db	1745	agcttcacataagagttccttatgtatttcaatggaacattcaaaagcccaatgatttgg	1804
OY	1784	cgaagcagag 1792	
Db	1805	caaatcagag 1813	
RESULT 3			
AAD09399			
ID	AAD09399 standard; cDNA; 1839 BP.		
XX	AAD09399;		
AC			
XX	10-SEP-2001 (first entry)		
DT			
XX	Vigna unguiculata neoxanthin cleavage enzyme, CPRD65 cDNA.		
DE			
XX			
KM	Cowpea; neoxanthin cleavage enzyme; abscisic acid; ABA; herbicide;		
KM	stress tolerance; transgenic plant; plant breeding; antisense-therapy;		
KM	plant growth protectant; CowPea Responsive to Dehydration; CPRD65; ss.		
XX	Vigna unguiculata.		
OS			
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..1839	
FT		/+tag- a	
FT		/product- "Vigna unguiculata CPRD65 protein"	
XX	EP116794-A2.		
PN	18-JUL-2001.		
PD			
XX	11-JAN-2001; 2001EP-0300218.		
PF			
XX	13-JAN-2000; 2000JP-0010056.		
PR	11-JAN-2001; 2001JP-0003476.		

XX	(RIKE) RIKEN KK.
PA	
XX	Iuchi S, Kobayashi M, Shinozaki K;
PI	
DR	WPI: 2001-400081/43.
XX	P-PSDB: AAE04787.
DR	
XX	A DNA encoding a protein with a neoxanthin cleavage activity for
PT	- producing transgenic plants with improved or decreased stress tolerance
PR	
XX	Claim 3; Page 53-56; 101pp; English.
PS	
XX	
CC	The invention relates to neoxanthin cleavage enzymes and their
CC	corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC	role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC	Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC	plant when expressed in a plant cell. The invention also relates to
CC	methods for increasing or decreasing stress tolerance in a plant by
CC	introducing the DNA into the plant, and a transgenic plant into which a
CC	neoxanthin cleavage enzyme is introduced. The improvement of stress
CC	tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC	cleavage enzyme genes are useful for producing transgenic plants. An arid
CC	land can be improved by growing transformant weed for several years and
CC	then removing the weed by specifically lowering stress tolerance in the
CC	weed by inducing an inducible promoter. The present cDNA sequence encodes
CC	Vigna unguiculata neoxanthin cleavage enzyme, CPRD5 (Cowpea Responsive
CC	to Dehydration) protein. CPRD5 gene is isolated from cowpea plant.
XX	
SQ	Sequence 1839 BP; 447 A; 595 C; 442 G; 355 T; 0 other:
<hr/>	
Query Match	44.4%; Score 798.4; DB 22; Length 1839;
Best Local Similarity	70.4%; Pred. No. 6,7e+248;
Matches 1098; Conservative	0; Mismatches 456; Indels 6; Gaps 2
OY	233 ccaaaagatccacaacttaacagatgatgtttccagagagcggcgycagcttgg 292
Db	
	281 ccaaccaaccattacctcaaaaatggaacttcttcagaagaagcgcgtcacgccttg 340
OY	293 acgcgagcgagagggttctctctgtcagccacgagaaagtacacccgcttctaagcgctg 352
Db	
OY	341 acctgtgcgaacgcgcgtcgtctcgcacagagcgaacaccgcgtcccccaaacgycgg 400
OY	353 atccctagtttcaagtcgcgcgcgaaatattgctccgcgttgatbaaaagcccgcgcgcgta 412
Db	
OY	401 acccgagagggtcccaaacgcgcgcggaacttcgcgcgcgttcgcgagacatgcgcgcgatcaa 460
OY	413 attcttcggttgcggaanaacttcccgatctccatcaaaagagtgatgtgcgcaacgag 472
Db	
OY	461 gactccgcgttgtcggaaaaatccccaatgatgatgacgcgtgtcatgcgycgcaacggtg 520
OY	473 ctaaccaccttacagagcgcgtgtacagaggtcacaccacttcttcgacgagagacggtatgttc 532
Db	
OY	521 caatlcoggtctacagcctgtgcygcgcgcacacacttcttcgacgagcgacgtatgtcc 580
OY	533 acgcgcgtcaaatltcgaacgcggttcaagctagcttagcgttcgcgcggtttactcagactaac 592
Db	
OY	581 acgcgcgtgaagttcacgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgttcacacgcgagcagc 640
OY	593 gggttgcgaagaagtcgaatgggtgcgcgcggtttccccaagaacatcgcgttgagcttc 652
Db	
OY	641 gtctctgcgagagaagaattctcagacgcgcgcgcggtgttcccgaaaggacacggygagctcc 700
OY	653 acgcgcacacccggtattgcgcgactcatgcttatctacgcgcgaagctgacgcggtatag 712
Db	
OY	701 acggcacactcgcgatacggcgcgctctctctcttcatacgcgcgcgcgtctcttcgggctcg 760
OY	713 tcgacgcgcgacacaggaacgcggtgtagctaacgcgcggttgcgttatcttaattcaatggcgcgt 772
Db	
OY	761 ttgatgtggtgccccaaggcatggcgctgtgcgaagcgcgcgtctcgtctactcaactcaacaacacc 820

QY	773	taaggttatgtctgagggatctgatttaaccttaccagaattccagatctccacatgagat	832
Db	821	ctctggtcccaatgctccgaagaacgatacttaacctaccacagctgagaataccccccttaacggtcagt	880
QY	833	taaaaacccgtgtgtctggtcttcgcatcttctgatactgagacaattagaaatccacaatgatactgcccac	892
Db	881	taaccacacgtgtgctggttaagacttcaacacgggcagctcgaattcaactcaacatgatactgcccacc	940
QY	893	cggaaatctgaccccggaataccggtgtgaactcttcgtctttaagctaaagacgttcgtttcaaaagc	952
Db	941	cgaataactctgagcccccgtctgacacgtgacacgtccacacgctccagctacagctacagcttcaagaagc	1000
QY	953	cttaacccaataatactccgatactctccatctcccggaacggaacataacacacggaactcgagatct	1012
Db	1001	cttaactcccaagatcttccgttctctcccccgaacggtcgatcaagctcccccgaactgagaatctcc	1060
QY	1013	agctgttgacagcccaacgatactgatacgaattctcgcgattacagaagaactctgctgcgtac	1072
Db	1061	ccccgaaagagccccaacatgatactgatacgaattctcgacataaagagaattctgctgcgtacc	1120
QY	1073	ctgagccagcaagctcgttttccaagctctgcccgaagaatgataccggtgtggtctcctcggtgtgt	1132
Db	1121	ctcgacccagcaggtgtggtcttcaaacatacgggaatgatacctcggcggtgtcccccggtgtct	1180
QY	1133	acgcaacgaagaacaaagttctgcgaagaattctggagattctgaacaaataacacgcgaaagaattatcga	1192
Db	1181	acgcaacgaagaacaaacotcaacgttctgtgattctctgacacagaatgctgaagagcagcgaatgt	1240
QY	1193	acataaagtgtgatactgatactgctccagaattctctctgcttccatctctcgagaacgtctggagaag	1252
Db	1241	cgaatcggtgtgatactgcagacgcgcgcggaattgtttctgtcttccacactctctgaaacgcgttggag	1300
QY	1253	agcccaagaacaaagaatctgaatctgtctgtatagaggtctccgtgtatctgactccacacgaatactat	1312
Db	1301	agccccgaacaaacgggaggtgtgtgtgattggttccctgcgaacgccccctgctggaactccatctt	1360
QY	1313	tcaacgaagctctgcaggaatactcaagaagtgtccctgtctgnaaatacccgctctgaaatctcaaaa	1372
Db	1361	tcaacggaatctcgagagagatctttaaagaacgctgtctgcagaagataaagctctgaaactctgagga	1420
QY	1373	ccggttgataactctgcgcgcgtccgatacatctccaacgaagaatacaacagtcacacctcgagag	1432
Db	1421	ccggtcaagctccatctgcgcgcgcccatatctccgaacgcgt--aaacagttgaacctggaag	1477
QY	1433	cagggaatgtgtcaacacagaatacatgtctgcgcgtataaacaacaaatctgcttactgtgttttag	1492
Db	1478	ccggtcatgtgtgaacacagaatacaagaactctggaagaagacccagttcctggtatctgtctgtgt	1537
QY	1493	ccgggacccgtgtgcttaagttctcaagttctcgctcaagatttgatactcaactctgtgagaagtta	1552
Db	1538	ccggagcccccgtgcccacaagttctcggtgtctgtgcgaagaattgatctgtgagtggtgagagatga	1597
QY	1553	agaaacactcttcaacgcgataacacgtcttaacggaggaagacgtctgtcttcccccggagaag	1612
Db	1598	agaagaatcatgtatctgagaagaagataagttctgtgtggagagacgtctgtttcttctccc--aacg	1654
QY	1613	gaggaagaagaacgcgaagaatataccctctgttctgcgttcacgaagagaagacatgagaat	1672
Db	1655	gccaacaaagaagaacgatactgtgttatattcttgcatctgtgcacagaagagaagaatgagaat	1714
QY	1673	cgaggttaacgatagtttaaagccgttagtttaggtttgaagcaaacggtttaaacttcgt	1732
Db	1715	ccggagctcgacgaatgtgtgaatgcccnaaattttaagaatctcgaagcttccatcaaacctcccc	1774
QY	1733	caagaggtctccgtatacgaatttcaacgtgtatacatcgcgaacccgatctgttcggaagcag	1792
Db	1775	ctcgtgtccctcaatcggtttcatcatgagaacttcatcattccaagaagatttgaggaacaag	1834

AC	AA009394,
XX	
DT	10-SEP-2001 (first entry)
XX	
DE	Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED1 cDNA.
XX	
KW	Neoxanthin cleavage enzyme; AtNCED1; abscisic acid; ABA; herbicide;
KW	stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW	plant growth protectant; ss.
XX	
OS	Arabidopsis thaliana.
XX	
FH	Key Location/Qualifiers
FT	CDS 1..1752
FT	/*tag= a
FT	/product= "Arabidopsis thaliana AtNCED1 protein"
XX	
PN	EP116794-A2.
XX	
PD	18-JUL-2001.
XX	
PF	11-JAN-2001; 2001EP-0300218.
XX	
PR	13-JAN-2000; 2000JP-0010056.
PR	11-JAN-2001; 2001JP-0003476.
XX	
PA	(RIKE) RIKEN KK.
XX	
PI	Iuchi S, Kobayashi M, Shinozaki K;
XX	
DR	WPI: 2001-400081/43.
DR	P-FSDb: AAE04782.
XX	
PT	A DNA encoding a protein with a neoxanthin cleavage activity for
PT	producing transgenic plants with improved or decreased stress tolerance
PT	-
XX	
PS	Claim 3; Page 18-22; 101pp; English.
XX	
CC	The invention relates to neoxanthin cleavage enzymes and their
CC	corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC	role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC	Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC	plant when expressed in a plant cell. The invention also relates to
CC	methods for increasing or decreasing stress tolerance in a plant by
CC	introducing the DNA into the plant, and a transgenic plant into which a
CC	neoxanthin cleavage enzyme is introduced. The improvement of stress
CC	tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC	cleavage enzyme genes are useful for producing transgenic plants. An arid
CC	land can be improved by growing transformant weed for several years and
CC	then removing the weed by specifically lowering stress tolerance in the
CC	weed by inducing an inducible promoter. The present cDNA sequence encodes
CC	Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED1 protein.
CC	The AtNCED1 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC	library using a cDNA of the CPRD5 (cowpea Responsive to Dehydration)
CC	gene isolated from cowpea plant as a probe.
XX	
SQ	Sequence 1752 BP; 484 A; 396 C; 428 G; 444 T; 0 other;
XX	
Query Match	41.0%; Score 738.2; DB 22; Length 1752;
Best Local Similarity	68.2%; Pred. No. 2.e-228;
Matches 1063; Conservative	0; Mismatches 478; Indels 18; Gaps 2.
OY	244 aaactataaacatgatattgtcttcacagagcgcgccgcttgcgaacggcgag 303
DB	193aatctcttcgccctaacaatcttcagaaggcgcgcatgtcgatcgaacggcgtag 252
OY	304 gattctcttgtaagcaacgagaaagtacaccgcgcttacctctaaaagcgctatcttatgttt 363
DB	253 cgttgatattaatcacacagcaagaattctccaacttcccacaaacgcgctatccacgttgtt 312
OY	364 cgatccgccgggaattttctctccgtlgaatgaacagccctccgcgcgtatattccggt 423

Db	313	caagatcgcggaattctatccccggaacggaattctccgtccgcggaaccacacgcgc	372
Oy	424	gtcggaaaacttcccgatcttcacaaagagtgatgtgcgaacgagactaaccaatt	483
Db	373	gaaggaacaacccctgactcattgacggtgttatatccgttaacggcgcgatccggtg	432
Oy	484	caagagccggtgacaggttcaaccctcttcgcagcagacgctatggttcaocgcgtcaa	543
Db	433	tttagcccaaacgcttgcgcacattatctgcagcagacggaatgttcaocgagttaaa	492
Oy	544	ttcgacaacggttaagtaactgaactgcttcggtttacttcagactcaacgggtttga	603
Db	493	ataaccaacggttcaagtaactgaactgaactgcggtttcaaaaaacggagatgttga	552
Oy	604	gaacgtcaatattgggtcgacacgggttttcccaaaagccatcggtgagattcaocgcacac	663
Db	553	gaaaaacgattgttggtcgacacagtttcccgaaagacatcggcgagcttcaocgttaccg	612
Oy	664	ggtatgcccgcactcatgtctattctacgcgcagagctgcagccggtatagtgcacccgycga	723
Db	613	ggaatcgacggtttgatgtcgtttttaacgcagcgggtcttggtgtcgtgatacaacaaca	672
Oy	724	caacggaacccggtgtgaactaaccccggtttggtctatttcaatggtccggttatgtctatg	783
Db	673	aacggtcgtcggatgatacagaacgcgcggttttggtttactttaaacaocggttttgaactatg	732
Oy	784	tcggagagatgtatttacttcaactcaagtttcagatctacaccccaatggagatttaaaacgctt	843
Db	733	tcagaaagacgatttcaactcgttaccatataaaatattcaactcaaacccggagatcccaacgctt	792
Oy	844	gttcggttccgattttgtatgtgacaattagaattcacacatgattgtccacccggaagtgcac	903
Db	793	ggaacttaacgatttcgcagcgttcagtaaaattccgaatgatagctcaacggaactgac	852
Oy	904	ccggaatccggtgtgaactcttcgcgttttaagctaacgacgtcgtttcaaaagccttaactaaa	963
Db	853	ccggtttagaagagagcttccagcgtttagaagctgaactacgacgttcgttaagaataaacttaactgtgaaa	912
Oy	964	tacttccgaattctcaacccggcgacgaactcaactcaacccgggaacgctgcgaattgattatcag	1023
Db	913	tacttccgaattctcgcgcagacgcgcgttaaatatcgcggaaattgagatcccgcttcgaaact	972
Oy	1024	ccaacgacatgacacgatttcgcgagattacagagaacatcgcgtcgaactgtgacacgaaca	1083
Db	973	ccgagacgtgatttcaacgatttcgcgtataacagagaaatttgggtgatttcctgatcaacaaca	1032
Oy	1084	gtcgttttccaagctcgcgcgagatgataccgcggtgtgggtctccggtgtgtgtttagacaagaac	1143
Db	1033	gtcgtgttcaacgctcgcgcgagatgatttccgcgttaaaatcccggtttgttttcgcagcgaagaa	1092
Oy	1144	aaggttcgcaaaagattccggaattttagaaacaaacagccgaagatttcaacgcgaacttaaggg	1203
Db	1093	aaggtttcccgattcgtgggataaagcccaaggaagcgcacagaactcttcagataactcgg	1152
Oy	1204	attgatgtcccaagattgtctctgtcttcacatctcttgnaaocgcttgggaagagccaganaaca	1263
Db	1153	gtgaactctccggagacgcttctgttttcatctcttgaatgacatgaggaatcgcgcggagacg	1212
Oy	1264	gatgaagtctgtcgtgataaggtctcgttatgactccacacgaactcaatttttaacgagctc	1322
Db	1213	gagagagattgtggtgataccgagatcgtgtatgtcgcgcggcgagatcaactcttcaacagagaga	1272
Oy	1324	gacgaagaattcccaagagttgtccgttgcgaatccgcgtgaatctcaaaaacccggtgaatca	1383
Db	1273	gacgagagcttgaagaagcgttttgcggagatacagagataaacctccgaacacgtataaac	1332
Oy	1384	actcgcgctccgaatcacatcccaacgaagatcaacaagaagtcaactcgaagccagagatgttc	1443
Db	1333	acggtcgtctcgtgtgtgtgttaacagagat-----gtaaatttagagattgtgatgttc	1386
Oy	1444	aacggaacatgtcgcgcgtlaaaacacaattcgtcttactgtgctttagcgcagccggttgg	1503

Db	1387	aacgcgaaccgggttgagagaagaaaaaccccggttcgcggttttgcgtatgcttalccttgg	1446
Qy	1504	ctctaaagcttcaggaattcgcgttaaagltgatactcaactcgtgagaagttaaagaacatctt	1563
Db	1447	ccaaagaattccggttcgctaaggttcgatactcttgcacccggttgagatgaaaaatatatt	1506
Qy	1564	tacgcgcgataacggttcaaggagagagcctctgtttctcccggt-----agaa	1611
Db	1507	tacgcgcggttgaaataatcagcgcgcgaacccgtttcttcgcgcgcgaactccggttaagc	1566
Qy	1612	ggagagagaggaagaagaagatactcctcgttctgcgttcacagcaagaagacatgaaa	1671
Db	1567	gaagaataatgaagaatgaagcgggttatattatcttgtaacggttcatacgagaagaacaagaca	1626
Qy	1672	tccgaggttacagatagttaacgcgcgtttagcttaagaggttgaagcaacggtttaaacttcg	1731
Db	1627	tcaagcgttcacgattataatcagctcgtttaattaaagcttgaagcgttaagcttaactaacg	1686
Qy	1732	tcaagaggttcggtacggaattcaacggtatcatcgcgaagccgagatgattggcgaagca	1790
Db	1687	tctagagtaacggtatcggtttcatatgacacatttgcgatactcgaatgaactcgttgaatca	1745
RESULT 5			
AAd09400			
ID	AAd09400	standard; CDNA; 1815 BP.	
XX			
AC	AAd09400;		
XX			
DT	10-SEP-2001	(first entry)	
DE			
XX			
KM	Zea mays neoxanthin cleavage enzyme, VP14 cDNA.		
KM	Maize; neoxanthin cleavage enzyme; VP14; abscisic acid; ABA; herbicide;		
KM	stress tolerance; transgenic plant; plant breeding; antisense-therapy;		
KM	plant growth protectant; Ss.		
XX			
OS	Zea mays.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..1815	
FT		/*tag= a	
FT		/product= "Zea mays VP14 protein"	
XX			
PN	EP1116794-A2.		
PD	18-JUL-2001.		
XX			
PF	11-JAN-2001; 2001EP-0300218.		
XX			
PR	13-JAN-2000; 2000JP-0010056.		
PR	11-JAN-2001; 2001JP-0003476.		
XX			
PA	(RIKE) RIKEN KK.		
PI			
PI	Iuchi S, Kobayashi M, Shinozaki K;		
XX			
DR	WPI; 2001-400081/43.		
DR	P-PSDB; ABE04788.		
XX			
PT	A DNA encoding a protein with a neoxanthin cleavage activity for		
PT	producing transgenic plants with improved or decreased stress tolerance		
XX			
PS	Claim 3; Page 60-64; 101pp; English.		
XX			
CC	The invention relates to neoxanthin cleavage enzymes and their		
CC	corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key		
CC	role in endogenous abscisic acid (ABA) biosynthesis under drought stress		
CC	Neoxanthin cleavage enzyme is used for improving stress tolerance in a		
CC	plant when expressed in a plant cell. The invention also relates to		
CC	methods for increasing or decreasing stress tolerance in a plant by		
CC	introducing the DNA into the plant, and a transgenic plant into which a		

PM EP1116794-A2.
 PD 18-JUL-2001.
 PX
 PX 11-JAN-2001; 2001EP-0300218.
 PF
 PX 13-JAN-2000; 2000JP-0010056.
 PR 11-JAN-2001; 2001JP-0003476.
 PR
 PX (RIKE) RIKEN KK.
 PA
 PX
 PX Iuchi S, Kobayashi M, Shinozaki K;
 PI
 PX WPI; 2001-400081/43.
 DR P-PSDB; AA04786.
 DR
 PX
 PT A DNA encoding a protein with a neoxanthin cleavage activity for
 PT producing transgenic plants with improved or decreased stress tolerance
 PT -
 PX
 PS Claim 3; Page 46-49; 10pp; English.

The invention relates to neoxanthin cleavage enzymes and their corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key role in endogenous abscisic acid (ABA) biosynthesis under drought stress. Neoxanthin cleavage enzyme is used for improving stress tolerance in a plant when expressed in a plant cell. The invention also relates to methods for increasing or decreasing stress tolerance in a plant by introducing the DNA into the plant, and a transgenic plant into which a neoxanthin cleavage enzyme is introduced. The improvement of stress tolerance in plants is useful, for example in plant breeding. Neoxanthin cleavage enzyme genes are useful for producing transgenic plants. An arid land can be improved by growing transformant weed for several years and then removing the weed by specifically lowering stress tolerance in the weed by inducing an inducible promoter. The present cDNA sequence encodes Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 protein. The AtNCED5 cDNA is obtained from an Arabidopsis plant-derived cDNA library using a cDNA of the CPD65 (Cowpea Responsive to Dehydration) gene isolated from cowpea plant as a probe.

Sequence 1734 BP; 464 A; 390 C; 441 G; 439 T; 0 other;

Query Match	29.3%	Score 528.2;	DB 22;	Length 1734;
Best Local Similarly	61.0%	Pred. No. 2.9e-160;		
Matches 920; Conservative	0;	Mismatches 563;	Indels 24;	Gaps 3;

OY	306	ttctctgtcagccacgagaaagcaacaccggtctctcctaataaagcgtgatctctatgtgttca	365
Db	240	tatcgtttcttcctatgttgagcgaaatcgcgcgcgtctctcctaataacgacgcgcacccgcgcgtgttca	299
OY	366	gatacccggaatatctgtctccggtgtgaattgaacacacccgcctccgcgcgtgatctctccgcgtgt	425
Db	300	attatcaggttaacttcgctccggtgttaattgtgcgtttcagaaacggtttagaagtggt	359
OY	426	cggaaacttccgattccatccaagaagtgatgtgtgcgaacgagctacaaccaacttca	485
Db	360	tgttgaagattctctctctgtctcaaaaggagatttaactcgttaacgctgtgcgaacctatgt	419
OY	486	cgaacgcgtgtgacaggtcacaccctctcttcgaagcgagacggtgatgtttacgcgcgtcaaat	545
Db	420	tcgcgcgtttacgcgacatactattttgacggtgacgaatgatattacgcgcgtttagat	479
OY	546	cgaacaacggt---taagctagtacgcgttgcgcgtttactactagaactaacccggtttgttca	602
Db	480	cggtttttatataccaggtttagtttacagctgcgcgtttacacttaaaacaaccgcgcttgttca	539
OY	603	ggaacgtcaattgtgttcgaccgcgttttccccaagaacatcgtgtgagtttaccaggccaac	662
Db	540	agaacacgcgcgtctgtgacatcgcgtttctccctaacaacatcgcgcgaagtttcaagggcaatc	599
OY	663	cggattttccgcgactcatgtctattctaacgcacagacgtgcagaccggtatgttgcaccgcgc	722

D	b	600	cggctctagctgcactgcgtctctcttcaacgctcgaagcttggatcggttccatgctggaacgggac	655
Q	y	723	acacgcgaaccggctgtagcttaacgcgcggcttggtgtctatttcaattcgaatggccggttatgtgctat	782
D	b	660	acgtgacatgggcgttagcttaacgcgcggcttggtgtctatttcaattcgaatgggttatattgacct	719
Q	y	783	gtcggagagatgatttaccttaccgaagtctgaagctacatctcccatctggaagtttaaaacgct	842
D	b	720	gtcgaagaatgatacttcccttaccgaatggaagatcgacggttcaaggagatccttggaacgat	779
Q	y	843	tgtgtgcgtctgattttgatgagacaattagatccacaatgagatctggccaccoccgaaagtctga	902
D	b	760	cggagcgttctggtatctgtagtaccgaagatttgactctcttcagtgtatgagtcgacatccaaagtga	839
Q	y	903	ccggagatccggctgtaactcttcgcttctaagctacgacgctcgctttcaaacgcttacctaa	962
D	b	840	cgcgaccacaaggagatcttccatctacactgtagatccaacgcttttggaaggaaaccttcatctga	899
Q	y	963	atacttcgcgattcttcacccggagaaactaatatcacccggacgctcgagatccatgctatga	10222
D	b	900	gtattcttaattcaacacccgctggcggaagaaacacgtgcgctggagatcaacgcttccctga	959
Q	y	1023	gccaaacgattgtagcagatttcgcgattacagaagaaccttcgttcgtctacatctgacacga	10822
D	b	960	accaacgattgatctatgatttctcgataaccggaatttcttcgttataccggtacatcgaca	1019
Q	y	1083	agctgtttcaagctgcgcgaagatgataccggcttggcttgcctcggtgttggttttcaacaaaga	11422
D	b	1020	aatgttatctcaaatcatctcgaanaagatctcggggcggtgatacccggttatacctaagttaaaga	1079
Q	y	1143	caaggtctgcaagatctcggatcttttagacaatacgcgcgaagatctcatcgaaataaagt	12022
D	b	1080	aaaatgtgcgagattttagtatttcttgaacgacgagatctgacccggttcggtatataatgt	11399
Q	y	1203	gattgagcttcgaattgcttctgcgttccatctcttgaaacgcttgggaagcgcaagaac	1262
D	b	1140	ggttgatgtacccgattctgttcttggttccatctcatgaaatgctgtggagagagaaacga	11999
Q	y	1263	agatgaag-----tcgtcgtgatagtgctcttctttagactccacgaatccatattt	1313
D	b	1200	agagggagacccagttatctgtcgttaacggtgatagtatgagccacccgacaagactct	1259
Q	y	1314	caacgagctctgaacgaagaatctcaagagtctcctgtcttgaaatccgcctcgatctccaac	1373
D	b	1260	tagtgaatcaagagaaacaaacccgggtttgattaaagttagactccggtttaaacaatgctac	1319
Q	y	1374	cgttgaaatcaactcgcgcgttcgcgatactcatctcaacgagaatcaacaaagtccaactcgaagc	1433
D	b	1320	aaaagaatcgaaccgctgaagttatcgttaac-----ggagtgaatttgaagc	1367
Q	y	1434	agggaatgttcaacagaaacatatgcctcggccgctlaaaacaaatcgcctacttgcttga	1493
D	b	1368	gggtacataaaacgctgttagctgtggccggaagaacccagttctgcttcatatgacatagc	1427
Q	y	1494	cgaagccgggtgcgttaagtctccgaagatttcgcttaagtttgttgcactactctggaagttaa	1553
D	b	1428	cgaatcccttgcccoaaatgacgtgtgcatcttggaaggttagatatcaaaaacggaaccggttct	1487
Q	y	1554	gaaacatcttaccgcgcgataaacgctttagcggagagagacgctcctgttctcccccggaaggt	1613
D	b	1488	agagtttaattagcagaccgcgcgcttgcgttggagaacgcgttcttctttagccggaggtga	1547
Q	y	1614	agggaaggaagaacgaagatacatctcctgttctcgttcaacgcgcgaagaagacatgnaatc	1673
D	b	1548	agggaagaagaacaaagattatgtatagtgttcttgagagacgcgaagaagaagacgaatctc	1607
Q	y	1674	ggagtgtacagatagttaaacgcgctttagcgtttagaggtgttggaacgcgtttaaactcgtct	1733
D	b	1608	ggagtttctgtgtgtcgaacgcgcgaagatatgaagcaagctcgcgcgcggtgtgccttgcgga	1667
Q	y	1734	aaggttctcgtacgatttcaacgtacatctcatcgcgagccgagatgttggcggaagcaggt	1793
D	b	1668	gaaggtacctattgtttccatgtgaacgctctcgtgaacgcgaatctcgttgaaggaacaggt	1727

XX WO200053724-A2.
XX 14-SEP-2000.
XX 09-MAR-2000; 2000WO-US06112.
XX 11-MAR-1999; 99US-0266513.
XX 18-AUG-1999; 99US-0149485.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX Wood M, McGrath A, Shenk MA, Glenn M;
XX WPI: 2000-579369/54.
XX New isolated polynucleotide encoding a plant transcription factor for
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,
XX having modified gene expression or modified activity of a polypeptide
XX -
XX Claim 1; Page 461; 747pp; English.
XX The present invention relates to novel plant transcription factors from
XX Eucalyptus grandis or Pinus radiata. The present sequence is the coding
XX sequence for one such transcription factor. The transcription factor may
XX be used to produce a plant having modified gene expression such as a
XX woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
XX mahogany species or to modify the activity of a polypeptide in a plant.
XX The transcription factors of the present invention are members from the
XX following families of regulatory proteins: bZIP, bZIP family of G-box
XX binding factors, basic helix-loop-helix zipper,
XX homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AR2
XX and ERBBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
XX and MYB.
XX Sequence 325 BP; 45 A; 128 C; 100 G; 52 T; 0 other;
XX
XX Query Match 9.5%; Score 171.8; DB 21; Length 325;
XX Best Local Similarity 71.2%; Pred. No. 4.6e-45;
XX Matches 227; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
XX
OY 451 ggaagtgtatgtcggaagagtaaccacttcacgagccggtgacaggtacacacttc 510
DB 3 ggggtgtactgctcggaacggtcccaaccgctccacgagcggtcggcggaacacacttg 62
OY 511 ttcgacgagagacggtatgttcaacgctcaaatcgaacacggttcaagctacagct 570
DB 63 ttcgacggtcgaagcgatgatccacgcttcggtcttcggtcgtctagtgagctaacgct 122
OY 571 tgcggttactcagactaacggtttgttcaggaacgtcaattggttcgacggtttc 630
DB 123 tgcggttactcagactaacggtttgttcaggaacgtcaattggttcgacggtttc 182
OY 631 cccaagacatcggtgacgagcgaacacggttatggtccgagctatgattctac 690
DB 183 cccaagacatcggtgacgagcgaacacggttatggtccgagctatgattctac 242
OY 691 gccacagactgcagccggtatagtcacccgagcagacggaacgggtgtagtaacggct 750
DB 243 gccgagcgtcttcgagcgttcgacacacgctaatgtagtggttcgagcgggac 302
OY 751 ttggtctatttcaatgggc 769
DB 303 ctctgttacttcgacggcc 321

AC AAC42989;
XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37611.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131049.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.

```
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140921.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144684.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 03-AUG-1999; 990S-0146389.
PR 04-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0148368.
PR 17-AUG-1999; 990S-0148175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.

PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151303.
PR 01-SEP-1999; 990S-0151330.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 14-OCT-1999; 990S-0159638.
PR 18-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0160989.
PR 25-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 25-OCT-1999; 990S-0161406.
PR 26-OCT-1999; 990S-0161359.
PR 26-OCT-1999; 990S-0161360.
PR 26-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161922.
PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

Query Match 9.4%; Score 169.6; DB 21; Length 1788;
Best Local Similarity 50.2%; Pred. No. 6,5e-44;
Matches 709; Conservative 0; Mismatches 634; Indels 69; Gaps 9;
```



```
QY 665 --gatttccgactcatgtctattctaacgccaagctgcagccggtatagtcacccgac 722
Db 633 gtacagtagctgtgagagctttaaaccgacgctagggttttaaccggaacgtataatccggt 692
QY 723 aacaggaacgggtgtagcttaacgcccgtttggtctattcaatggccggtatttggctat 782
Db 693 taacgcatctgttagtaataacaagcttagctttcttcagtaaccgctctcttgcctt 752
QY 783 gtccgagagatatttaccattcaagttcacatctcccaatggagattttaaaccgt 842
Db 753 aggtgaatctgttaaccctacgcccgcgtacgttaacccaatcagagatatttgaacag 812
QY 843 tggctgctcatttcttgatgagacaattagaaatcccaatgattgcccaccgaaagtcga 902
Db 813 cggacgagtagcatcttcgacgggaattagcgatgtagttagacagctatctccaaaccga 872
QY 903 ccggaatccggtgtaactcttcgctttaaagtcagcgtcttcaaaagccttaccataa 962
Db 873 tccaataaccggaagaattcttcgcttccggtacgctcgttcca---ccgtttttaac 929
QY 963 atacttcgattctcacgagcggaactaaatcacggaagtcgaga---ttcacctga 1019
Db 930 atattccggtttagattccgcgggaaaaaacaagaagacgctccgaatattctcgatgac 989
QY 1020 tcaagccaacgatlgaagcagatttcgcgattacagagaaacttcgctgcgtacccgaaca 1079
Db 990 gtctccgctgttctccatgacttcgcgacatcagaaacgctacgagatttccgacagat 1049
QY 1080 gcaagtcg-----ttccaagctccgagagatgataccggttggtctcgggtggt 1130
Db 1050 tcaagcttggcatgagagaaatgtgatttggttctcgaaagtggttctccggtctgg 1109
QY 1131 ttacgacaagaagaagtcgcaagaattcggaattttagaacaatccgcgcgaagatcatc 1190
Db 1110 tactgataacggaanaaacctcaaggcttgagtgatctcctaagtcgcggagagagatc 1169
QY 1191 gaacattgaagtgtatgtagtctccagatgtcttcgtctcaatccctcgaacgcttgga 1250
Db 1170 ggaagtagaagtggttcgaggttcctcgatctcaatataatcactacgtatgaagcttggga 1229
QY 1251 agagccagaacaacgataagtcgctgtagtaggttcctgtatgactccacagactcaat 1310
Db 1230 tgaagatgtagaaacagcgtctttagtcacccaatattatgcatgataacataac 1289
QY 1311 ttccaagagctcgaaggaatctcaagagtgctcgtctgaataacccgcgcgaactcaaa 1370
Db 1290 tttagagagagtgatcgtgttc---atgcttggtagagagtggaagatcgaatcgt 1346
QY 1371 aaccggtgaatcaactcgcgtccgatacatctccaacgaagatacaacagtcacactcga 1430
Db 1347 caccgggattgtgagaagctacgtactcgaagcagag-----aatctcga 1391
QY 1431 agcaggagtggtcaacagaacaatgctgcgcgttaaaccaaatctcgcttacttgctt 1490
Db 1392 ttccgctgtgatttaactccgagcgttctcggagagatgtagcaggttactgaacggcgat 1451
QY 1491 agcccgagcggtgcttaagctcagagattcgtaagttcactacatccatccagggagagt 1550
Db 1452 tggagatccgattgcgaagatctccggtcgtgtgtagagcttgatggtctaaaggagatcg 1511
QY 1551 taagaaacat-----ctttagcgcatataccgtttagcggagagaacccctct 1595
Db 1512 ggaatgattgtagcgtggtccgtagaagttagcgttcaagtgtagtcggcggaacggtt 1571
QY 1596 gtttctcccgagaga-----aggaaggagagagaagaagatacatatccctctg 1643
Db 1572 ttctgtagctgaggaatccgtgtaacatccgagagcggaagagagatgtagttagtgtagc 1631
QY 1644 ttctgtagcagagaagaacatggaatactcgaggttacagatagtagtaaccggttagctt 1703
Db 1632 gtatgttcaacgataagagatcgtggaatcgaaagttctgtgtgtagcgaacgaataatcgcc 1691
QY 1704 agaagttgaa-----gcaacggttaaaccttcgcgtcaagaggttccgtagcgaattcacg 1757
```

```
Db 1692 ggaagcttgaataatcgtcgcgcgcgtgaggttgcgggaagaggttccgtagcgaattccatg 1751
```

```
QY 1758 tactatcatcggaagccgatgatttggcgaagc 1789
Db 1752 gtatttgcgaaggaaagtgaccttaataagc 1783
```

RESULT 12

AAD09395 standard; cDNA; 1788 BP.

AAD09395;

10-SEP-2001 (first entry)

Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2 cDNA.

Neoxanthin cleavage enzyme; AtNCED2; abscisic acid; ABA; herbicide;

stress tolerance; transgenic plant; plant breeding; antisense-therapy;

plant growth protectant; ss.

Arabidopsis thaliana.

Key Location/Qualifiers

EP116794-A2.

18-JUL-2001.

11-JAN-2001; 2001EP-0300218.

13-JAN-2000; 2000JP-0010056.

PR 11-JAN-2001; 2001JP-0003476.

(RIKE) RIKEN KK.

Iuchi S, Kobayashi M, Shinozaki K;

WPI: 2001-400081/43.

P-PSDB; AAE04783.

A DNA encoding a protein with a neoxanthin cleavage activity for

producing transgenic plants with improved or decreased stress tolerance

Example 10; Page 25-29; 101pp; English.

The invention relates to neoxanthin cleavage enzymes and their

corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key

role in endogenous abscisic acid (ABA) biosynthesis under drought stress.

Neoxanthin cleavage enzyme is used for improving stress tolerance in a

plant when expressed in a plant cell. The invention also relates to

methods for increasing or decreasing stress tolerance in a plant by

introducing the DNA into the plant, and a transgenic plant into which a

neoxanthin cleavage enzyme is introduced. The improvement of stress

tolerance in plants is useful, for example in plant breeding. Neoxanthin

cleavage enzyme genes are useful for producing transgenic plants. An arid

land can be improved by growing transformant weed for several years and

then removing the weed by specifically lowering stress tolerance in the

weed by inducing an inducible promoter. The present cDNA sequence encodes

Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2 protein.

The AtNCED2 cDNA is obtained from an Arabidopsis plant-derived cDNA

library using a cDNA of the CPRD65 (Cowpea Responsive to Dehydration)

gene isolated from cowpea plant as a probe.

Sequence 1788 BP; 444 A; 444 C; 418 G; 482 T; 0 other;

Query Match 9.4%; Score 169.6; DB 22; Length 1788;

XX This invention relates to defence-related signalling genes isolated from
 CC the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage
 CC enzyme (NCE), an amino acid perase (AAP) and a glutamic acid rich
 CC protein (GRP). The signalling gene is useful for increasing the
 CC resistance of a plant to a pathogen such as fungus, virus, bacterium,
 CC nematode or insect (e.g. European corn borer), preferably
 CC *Sclerotinia* spp., *Phoma* spp., or *Phomopsis* spp., by stably incorporating a
 CC construct containing the gene into the genome of the plant. The gene is
 CC useful for regulating gene expression in a plant. In response to a
 CC stimulus such as infection with a pathogen, damage from a pathogen,
 CC hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,
 CC oxalic acid or expression of a gene encoding oxalic acid oxidase. The
 CC genes are also useful for stem-preferred regulation of gene expression in
 CC a plant. The genes are useful in agriculture, particularly in the
 CC breeding of crop plants with improved agronomic traits, for modifying
 CC abscisic acid (ABA) metabolism and for modifying amino acid transport and
 CC content in plants. The present sequence represents cDNA encoding the
 CC sunflower neoxanthin cleavage enzyme (NCE).

SO Sequence 1950 BP; 535 A; 456 C; 453 G; 506 T; 0 other;

Query Match 9.4%; Score 169.2; DB 22; Length 1950;
 Best Local Similarity 51.0%; Pred. NO. 9.2e-44;
 Matches 480; Conservative 0; Mismatches 453; Indels 9; Gaps 3;

QY 347 cggctgattcctagtgctcagatcgcgcggaatttgcctcgggtgaatgaagaacccgcgc 406
 DB 308 cagcttgatcccaaaacagcttctgtgataactttcaacggtgagacatccctccga 367
 QY 407 ggcgtatcttcgcgtgctgcgaaaactccgcattccatcaaaagagtgtatgcgca 466
 DB 368 ctgactgtgaagatcatcgagcgacactgcgcgaagtgtcgttcgaacgttacttcogta 427
 QY 467 acggaactaaccaactccacgagccggtgacaggtacaccacttctcgaggaacgta 526
 DB 428 atggtccgaaccgcgaattctcttcgcgagagaccatccactctcatgtgagtagca 487
 QY 527 tgggtcagcgcgtcaaatctgaaacggttcagtagtagtctgcgcggttactcaga 586
 DB 488 tgcctcagtcattctgatacccaatggaagaagcttgatgtagccgatacatcaaaa 547
 QY 587 ctacacggttgttcaggaacgtcaatgtgtgcacgcggtttcccccacaaacatcggtg 646
 DB 548 catacaaatatcaatagagaagaagcggtatcccatattccaaacgttgttttag 607
 QY 647 agcttcacgacacacgcggtatgcccgaatcatgtatctc--taacgcaagctgcag 703
 DB 608 gtttaattgtgtgactgctctcgtcagctgcagtcagtcacgtgcgcgcgcattttgg 667
 QY 704 ccggtatagtcgacccgcgcagacggaacgcgtgtgatacgaacgcggttggctatttca 763
 DB 668 ctgacaactttgacccccaagaaggtatgtgtcgaataccagtcagtcctttttg 727
 QY 764 atggcgcgttatgtgtcgtatgcgagagatgttaacttcaacaggttaagtaactccca 823
 DB 728 gcaacgaacttttgcctcgtggaagtcgagatcccatatgcgcgtcaactagcgccg 787
 QY 824 atggaatttaaaaacgcgtgtgtcgatttgaatgaacattagaatcacaatga 883
 DB 788 acggtgacatagtcacgcgtgcagctgagacttcgacgcaaaactattcatgacatga 847
 QY 884 ttgcccacccgaagaagtgcacccgaatcgggtgaactcttgccttaagctacgacgtcg 943
 DB 848 ccgcctacccaaaatgcatccaglaaagaagaagctttgtcctttgtaagcttcag 907
 QY 944 tttaaacgcttacttaaaatcactccgattccacccgaggaactaataccagcagc 1003
 DB 908 tcgcc---ccttcttaaccttttgcgtttcaacgaaagcggaagaaacacgacgtgtg 964
 QY 1004 tcgagatcagc---ttgatcagccaacgagtatgcagatcttcgagattcagagaact 1060

DB 965 tcccgatcttccaatgacgaagccgcgtcttctccacgacttcgcacatccacaaaact 1024
 QY 1061 tcgtcgtgtacccctgaccagaagctcgtttcaagctgcgcgagatgtagtcgcgtgtgt 1120
 DB 1025 acgcgatttcccccagatcccaaacgcggatgagacccaatggaatgctcgtgtggtgat 1084
 QY 1121 ctccggtgtttagcagacaagaagctgcgaagaatctgggattttagacaataacgcgc 1180
 DB 1085 ccccggttagcgcgagcgtcgtgaagaagtgctcgtcgtcgtggtgatccctcgttagcga 1144
 QY 1181 aagattcagacaactaagtgtgattgctcgaagatgcttgccttcacatctctga 1240
 DB 1145 aagcagctccagatgagatggtgttgaggttcgcgggttttaaatgtgatacatgcatca 1204
 QY 1241 acgcttggaagagccagaacaagatggaagtcgtcgtgatag 1282
 DB 1205 atcatggaagagagatggtgcgagatcagtcgtgtgtgtg 1246

RESULT 14

AAC57157
 ID AAC57157 standard; DNA; 491 BP.

AC AAC57157;
 DT 25-JAN-2001 (first entry)

DE Pinus radiata transcription factor DNA sequence #603.

KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; ERBBS; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB, ss.

OS Pinus radiata.

PN W0200053724-A2.

PD 14-SEP-2000.

PE 09-MAR-2000; 2000WO-US06112.

PR 11-MAR-1999; 99US-0266513.

PR 18-AUG-1999; 99US-0149485.

PA (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Wood M, McGrath A, Shenk MA, Glenn M;

XX WPI: 2000-579369/54.

PT New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide

PS Claim 1; Page 598; 747pp; English.

XX The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and ERBBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.

SQ Sequence 491 BP; 86 A; 143 C; 171 G; 91 T; 0 other;

Query Match 8.3%; Score 149.4; DB 21; Length 491;
Best Local Similarity 59.2%; Pred. No. 1.1e-37;
Matches 255; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 436 ccgattccatcaaaaggagttatgtgcgaacgagctaaaccattcacgagccggtg 495
DB 1 ccgagagcttagatagatggtttatgcgcgaatgscgatcccggttcaaacccgc 60
QY 496 acaagtcacacattcttcgaacgagatgtgttcacgcccgtcaaatctgaacacggt 555
DB 61 ggcgcacacattatttcgacgagcgatgatatacatgcgttcgacgtgagacacggt 120
QY 556 tcagctatctcgcgttcggttcctacgactaacggtttgttcaggaagatcatgt 615
DB 121 aagctatgtacagtctgcgcgttcacgagacggaaggctcgttagcgaagagcggtgc 180
QY 616 gtcgaccggttttccccaagcactgcgtgagcttcacgagccacacggtattgccga 675
DB 181 ggcgcgagcttttaccggaagcccatcggtgcacacccagcgagcggtgtgtgcgc 240
QY 676 ctcatgtattctacacgagagctgcagccggtatagtcgacccgagcagacgagcgt 735
DB 241 ctgcgtcgtcagatgtgcgcgcgggtctgcgtgcgttcacacacggaagggcgtgcgc 300
QY 726 gtagctaacgagcgtttgtctatctcaattcagcgcgttatgttcgttcgagagatgt 795
DB 301 gtgcctaatgtccggcggtgcgtcttcttaacgagcgtctgcgtcatgttcgagaagcat 360
QY 796 ttaccttaacgaatcagatcacctcccaatgtagatlttaaaaacggttcgttcgtcat 855
DB 361 ctcccgatgcgcgtcaggtgacggtgacggtgacggtcgttcgttcgacgagcgttcgtcat 420
QY 856 ttggtatgaca 866
DB 421 ttgcgagggaca 431

RESULT 15

AAAC57162 ID AAC57162 standard; DNA; 386 BP.
AAC57162; AC
XX 25-JAN-2001 (first entry)
DT
XX Pinus radiata transcription factor DNA sequence #608.
DE
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; ERBS; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
XX Pinus radiata.
XX
XX WO200053724-A2.
XX
XX 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US06112.
XX
XX 11-MAR-1999; 99US-0266513.
XX 18-AUG-1999; 99US-0149485.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Shenk MA, Glenn M;
XX
XX WPI; 2000-579369/54.

XX New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PS
XX Claim 1; Page 599; 747pp; English.
PS

CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper, LIM domain, AP2
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and ERBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.

SQ Sequence 386 BP; 89 A; 105 C; 106 G; 86 T; 0 other;

Query Match 8.0%; Score 143.4; DB 21; Length 386;
Best Local Similarity 70.0%; Pred. No. 8.4e-36;
Matches 208; Conservative 0; Mismatches 86; Indels 3; Gaps 1;

QY 1103 agatgctcgcggtgtgtctcgcgtgtgttcgacgaagaacaggttcgaagattcggga 1162
DB 1 aaatgatcagaagcggtctctccagttatccacaacaaagaaaggtccgcgcttcgggc 60
QY 1163 tttagacaatacgcgcgaagatctcgaacattatgattgtatgtctcgaattgtc 1222
DB 61 ttctcccaataatgtctctgcgagagtgagctgaaatgatcgaaggtcccgattgtc 120
QY 1223 tctgttccatctctcgaacgcttcgggaagcgaagaacagatgaagtcgtcgtatag 1282
DB 121 tctgttccatctctcgaacgcttcgggaagcgaagaacagatgaagtcgtcgtatag 177
QY 1283 gttcctgtatgactccacagagctcaatttcaagaagcttcgaagatctccaagagt 1342
DB 178 gttcctgtatgactccacagagctcaatttcaagaagcttcgaagatctccaagagt 237
QY 1343 tctgtcgtgaatccgctcgaatctcaaaacggtgaaatcgaactcgcgtccgatca 1399
DB 238 ttctgtcgaatccgctcgaatctcaaaacggtcgttcacacgagcgagatca 294

Search completed: July 24, 2002, 06:25:34
Job time: 6890 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 05:14:15 ; Search time 1694.89 Seconds
(without alignments)
14333.980 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800
Sequence: 1 atggctcttcacgagcaac.....tggcgaagcaggtctgtgta 1800

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hlc: *
9: gb_est1: *
10: gb_est2: *
11: gb_hlc: *
12: gb_gss: *
13: em_gss_hum: *
14: em_gss_inv: *
15: em_gss_pln: *
16: em_gss_vtc: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	425.6	23.6	720	10	BM412731 EST587047
2	417	23.2	805	10	BM408615 EST582942
3	416.2	23.1	781	12	BH549344 BGR053TF
4	382.8	21.3	643	9	AW933524 EST359283
5	375.6	20.9	787	10	BM408565 EST582892
6	344	19.1	592	10	BF113346 EST440936
7	332.4	18.5	566	10	BE461924 EST413439
8	323.2	18.0	553	10	BM085672 sa128a02.
9	318	17.7	559	10	BM085005 sa131a08.
10	316.6	17.6	520	10	BE432853 EST399478
11	313.8	17.4	742	12	BH559796 BOHR047TF
12	312.8	17.4	594	12	BH458011 BOHR02TF
13	309.4	17.2	547	10	BI974879 sa174b11.
14	308.6	17.1	547	10	BE437072 EST408190
15	303.4	16.9	617	10	BE458861 EST414153
16	300.4	16.7	546	10	BM084948 sa130b08.
17	295.8	16.4	618	9	AW443298 EST308228

18	294	16.3	564	10	BM085488 sa137d09.
19	292.8	16.3	627	10	BF051297 EST436472
20	289.2	16.1	509	10	BE451573 EST402461
21	288.8	16.0	495	10	BE434930 EST406008
22	276.2	15.3	490	10	BF050563 EST435721
23	275.4	15.3	657	9	AA556214 69 Lobiol
24	267.4	14.9	503	12	B27476 T9A21TR TAM
25	254.8	14.2	566	10	BE459895 EST415187
26	249.2	13.8	450	10	AW930245 EST340702
27	248	13.8	552	10	BG038692 NXSL_102.
28	245	13.6	778	12	BH579654 BOGUH/31R
29	236.8	13.2	600	12	BH458018 BOHR02TR
30	233.2	13.0	424	12	BH496327 BOHWT2TR
31	224.4	12.5	337	12	BH579641 BOGUH/3TF
32	220.4	12.2	410	10	BE525530 M26D7STM
33	217	12.1	446	10	BE471089 WHE0283_A
34	216.6	12.0	713	10	BI928174 EST548063
35	209.6	11.6	495	9	AU084510 AU084510
36	207	11.5	502	12	CNS00DKV
37	205.6	11.4	590	10	BG592987 EST491665
38	203.2	11.3	447	9	AM933245 EST359088
39	202.8	11.3	492	9	AU084760 AU084760
40	201.4	11.2	412	9	AV419581 AV419581
41	195.2	10.8	680	12	AQ969739 LRPJ65TF
42	185.8	10.3	549	10	BF425834 su16e06.Y
43	173.6	9.6	500	9	AM289745 NXNW005D0
44	164.6	9.1	678	12	BH593312 BOHWT19TF
45	152.6	8.5	267	10	BE529289 W74M11STM

ALIGNMENTS

RESULT 1
BM412731 720 bp mRNA linear EST 22-JAN-2002
LOCUS EST587047 tomato breaker fruit Lycopersicon esculentum cDNA clone
DEFINITION CLEGG60N24 5' end, mRNA sequence.

ACCESSION BM412731
VERSION BM412731
KEYWORDS
SOURCE
ORGANISM

Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE
1 (bases 1 to 720)
Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai

J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Ronning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished (2002)
Contact: CWGI

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

This clone is available through the Clemson University Genomics Institute

Seq primer: T3.

Location/Qualifiers

1..720
/organism="Lycopersicon esculentum"
/cultivar="T496"
/db_xref="taxon:4081"
/clone="CLEGG60N24"
/clone_lib="tomato breaker fruit"
/issue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSMCadapt; Site_1: EcoRI;
Site_2: XhoI; supplier: Boyce Thompson Institute;

Qy	445	atcaaaagagtgatgctgacacggaactaacaccactltcaegagccggtgacaggtac	504
Db	245	gTTCAAGGCGTTTACGTTACGAAACCGAGACTAACCCCTCTTTTGGAAACCAACCCCGCAGAC	304
Qy	505	caactcttcagacgagacgcatgattgttcaagccgctcaaatcctgaacacaggttcagctagc	564
Db	305	CATTCTTCGACGGCGAGCGGTATGGTTTACGCCGCTTCATTTCATAAATGGGTGGCTACT	364
Qy	565	taagcttcgaggtttactcagaactaacccggtttgttccagaaagtcgaattgctgcagccg	624
Db	365	TACGCTTCGCCGTTTCATGAAACAGAGAGCGCTGTTCAGAAAGAAAGCTTTGGGTGGCCT	424
Qy	625	gtttcccccagaagcactcgtgtgagctltcaagccacacacggtatgtcccgactaigtcta	684
Db	425	GTTTCCCTTAAAGCCATTTGGTGAATTAATGATGGTCACTGTGGAATGCAAGGCTTATGCTG	484
Qy	685	ttctacagccagagctcgacagccggtatagtcgaccccgacacagagacccggtgtagctaac	744
Db	485	TTTACGCTCGGGGGCTCTTCGGACTTGTGTATCAACAGTAAAGAACTGGCTTCCAAAC	544
Qy	745	gccggtttgctcatlttcaatgcccggtatgtgatatgctcgagagatattactctac	804
Db	545	GCCGGTTAGTCTATTTCATATACCGGATTACTTGTGTATGTCTGAGATGATTGGCTTAC	604
Qy	805	caagtcagatcactcccaatgagatltaaaacccgtgtgctcggttcgatttgatgta	864
Db	605	CATGTAAGAGTAAACCCACCGCGCATCTTAAACAGAGGCGTGCATTTGATTCGAGCGC	664
Qy	865	caatagaatccacatgatgtgcccaccccgaaagctcgacccggaatccggtgttaactctc	924
Db	665	CAGCTAAATCCACCATGATATACCTACCCCAAGCTCGACCCAGTTCCGCTGAGCTATT	724
Qy	925	gcttaagcttcagacgttcgtttcaagcccttcaaatatcttcgattctca	978
Db	725	GCTCTACTACGATGTGATTCAGAAAGCCATACCTCCAGTACTTTCAGATTTC	778
RESULT	6		
BF113346			
LOCUS	BF113346	592 bp	mRNA
DEFINITION	EST440936	tomato breaker fruit	Lycopersicon esculentum
ACCESSION	BF113346		CDNA clone
VERSION	BF113346.1	GI:10943036	
KEYWORDS	EST.		
SOURCE		tomato.	
ORGANISM		Lycopersicon esculentum	
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.	
AUTHORS		Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Rongning,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.	
TITLE		Generation of ESTs from tomato fruit tissue, breaker stage	
JOURNAL		Unpublished (2000)	
COMMENT		Contact: CUGI	
COMMENT		Clemson University Genomics Institute	
COMMENT		Clemson University	
COMMENT		100 Jordan Hall, Clemson, SC 29634, USA	
COMMENT		Email: http://www.genome.clemson.edu/orders/index.html .	
FEATURES		Location/Qualifiers	
SOURCE		1..592	
		/organism="Lycopersicon esculentum"	
		/cultivar="TA96"	
		/db_xref="taxon:4081"	
		/clone="CLEG44B9"	
		/clone_lib="tomato breaker fruit"	
		/tissue_type="Pericarp"	
		/dev_stage="breaker"	
		/lab_host="SOLR"	

/note="Vector: pBluescriptSMCadapt, Site1: EORI; Site2: XhoI; supplier: Boyce Thompson Institute; sequencing: The Institute for Genomic Research. Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of fruit). Fruits were cut in half and the seeds and locules were discarded prior to freezing the pericarp."

BASE COUNT	194 a	106 c	131 g	161 t
ORIGIN				
Query Match	19.1%	Score 344;	DB 10;	Length 592;
Best Local Similarity	73.8%;	Pred. No. 1.5e-91;		
Matches 437; Conservative	0;	Mismatches 155;	Indels 0;	Gaps 0;
OY	893	cgaaagtcgaccgcgaatccggtgacatcttgcgtttaagccagcagcgttcgcttaaacg	952	
DB	1	CAAAAGCTGACCCCAATGTTCCGGTGAGCTATTGGCTTTAGTACGATGATGATTCAGAAAC	60	
OY	953	cttaacctaaatactctccgattctcaccgcggaactaaataccgcgagcgtcgatc	1012	
DB	61	CATTACTCAAGTACTTCAGATTTTCAAAAANAATGGGAAAANAATCAATGATGTTGAATTC	120	
OY	1013	agcctgatacgcgaacagatgatgcacgatttcgcgattacagaacttcgtcgttac	1072	
DB	121	CAGTTGAAGAGCCCAACAATGATGATGATTTGCAATTAAGTACGAACTTCGTCATATTC	180	
OY	1073	ctgaccagcgaatctgttttcaagctgcgcggaatgatccggtgtgtctccggtgttt	1132	
DB	181	CTGATCAACAAATGCTTTTCAAGATGTCGAATATGATCCGTGAGGTTCCACCGTGCTTT	240	
OY	1133	acgacaagacagaagtcgcgaagatccggtatttagacaatacgcgcgaagatlaacga	1192	
DB	241	ACGACAGAAGACAAAGTTTCCGATTTGGTATTCTGGATTAAGTACCGGAAGAATGGGTCG	300	
OY	1193	acattaatgatgatgatgctccagatltgctctgtcttcacatccttggaaagcttggaa	1252	
DB	301	ATTGAAATGGGTTGAAGTACCTGATTTGTTCTGTTCCTCCACCTGGAATGCTTGGGAAG	360	
OY	1253	agccgagaacagatgaatgctgcggtatgggttcctgttatgtatccacagatcaatt	1312	
DB	361	AAGCAAGAAACAATGAAATCGTTGAATGGTTCATGATGACACCAACGACTCCATTTT	420	
OY	1313	tcaacgagctcaacgagaatcacaagatgtctctgtcctaataccgcctgaatcca	1372	
DB	421	TCAATGAATGTCATGAAAGGCTTAAGAGTGTTTTATCCAAATTCGCTCAATTGGAAAA	480	
OY	1373	ccggtgatacaactgcgcgtccgcatcatctccaacggaatcaacaactcaactcga	1432	
DB	481	CAGGGAATCAACAAGAAATATCCATATGCAAAACCCGGATGACAAAGATTTAAGG	540	
OY	1433	caggagatggtcaacagaaacatgctgcgcgctgaanaacaaatcgcttact	1484	
DB	541	CTGGATGGTGAACCGAAACAATCTCGGAAGGAAAAACAGATGATGCTTATTT	592	
RESULT 7				
LOCUS	BE461924	566 bp	mRNA	linear
DEFINITION	EST141343 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA			
ACCESSION	BE461924			
VERSION	BE461924.1	GI:9506322		
KEYWORDS	EST.			
SOURCE	tomato.			
ORGANISM	Lycopersicon esculentum			
LOCUS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
DEFINITION	Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;			
ACCESSION	Asteridae; euasterids I; Solanales; Solanaceae; Solanum;			
VERSION	Lycopersicon.			
REFERENCE	1 (bases 1 to 566)			
AUTHORS	Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,			
	Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Roming,C.M.,			

QY	882	gagggccaccgcgaagctcgaccccggaatcggtagaactcttcgctttaaagttacgaagt	941
Db	61	GATGCTCACCCTCAACTCGACCCCGTACCAACAACCTCCACGCCCTCTCTACGAGCT	120
QY	942	cgttcaaacgcttactctaataacttcgaattctccaccggaacggaactaaatcacgga	1001
Db	121	CGTCCAGAGACCAATATCTCAAGTACTTCCGCTTCTCCCGACGGCTCAATCCCGGA	180
QY	1002	cgtcgaagttacgtttagtcggccaagatgatgcagatttcgattacagaagaactt	1061
Db	181	CGTGAATTTCCCTTGAAGAACCACCATGATGACATTTCCCATCACGGAACCTT	240
QY	1062	cgctgcgtacacctgacacgaagaatcgcttttcaagctgcggaagatcgcgcyggtgc	1121
Db	241	CGTGTGTCCTCCCGACGACGAGTGTCTTCAAGTTATCAGAAATGATCACCGGAGCTC	300
QY	1122	tcgggtggttttaacgaagaacaaagtcgcgaagatctgggattttaagcaaatacgcga	1181
Db	301	CCCCGTTTCTACGCAAGAACAAGGTTTCCAGATTCCGGATTCTCGACAGAAATGCTAA	360
QY	1182	agatcacatcgacacatlaagtgtatgattgcctccagattgctctgcctccatctggaa	1241
Db	361	AGATGCAATGATATNGAATGTGATGATGATGCTCCGAGTCTTCTTTCACCTCTGGAA	420
QY	1242	cgcttgggaagagccgaagaacagatgaaagtcgctgtagataggttccttatactcaccc	1301
Db	421	CGCCTGGGAGGACGCGGAAATGATGATGATGCTGTTGTCATCGGTTCTCCTCATGACCCGCG	480
QY	1302	agactcaatttttaacagagctctgcagaaatctcaaggtctccgtctgaatccgct	1361
Db	481	GGACTCCATTTTCAACGATCGAAGAGAAGTTTGAAGAGCATCTTGTGAGATTAAGCT	540
QY	1362	gaatccaaaac	1373
Db	541	GAATTTGAAGAC	552
RESULT	9		
BM085005			
LOCUS	559 bp	mrna	linear
DEFINITION	sa31808.y1 Gm-c1066 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-c1066-4240.5, similar to TR:024023 024023 NEXANTHIN CLEAVAGE ENZYME: , mRNA sequence.		
ACCESSION	BM085005		
VERSION	BM085005.1	GI:16995633	
KEYWORDS	EST.		
SOURCE	soybean.		
ORGANISM	Glycine max		
REFERENCE	1 (bases 1 to 559)		
AUTHORS	Shoemaker, R., Keim, P., Vodkin, L., Erpelidg, J., Corvelli, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, J., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ratter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.		
TITLE	Public Soybean EST Project		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact: c@resgen.com web site: www.resgen.com		
	High quality sequence stop: 421.		

FEATURES
source

Location/Qualifiers
1. .559

```

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl066-4240"
/clone_1ib="Gm-cl066"
/tissue_type="leaf and shoot tip, salt stressed, 2 week
old seedling"

```

```
/lab_host="DH10B"
```

seedling from the cultivar Williams. The 2 week old seedlings were salt stressed in a solution of 500mM NaCl for 3 days prior to harvesting. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

BASE COUNT	132 a	206 c	105 g	116 t
ORIGIN				

Query Match	17.7%;	Score 318;	DB 10;	Length 559;
Best Local Similarity	-73.1%;	Pred. No. 8.7e-84;		
Matches 408;	Conservative 0;	Mismatches 150;	Indels 0;	Gaps 0;

OY	628	ttcccaaaagccatcgggttgaagcttcaagcgccacaaccggatttgcggaactaatgcatc	687
Db	2	TTCCCCAAGGCCATCGGTGAACCTCCAGGGCCATCCGGCATCGCCCGCTCCTCTCTTC	61
OY	688	tacgcagagctcgaagccggatatagttcgacccggcacaacggaaccgggttagtctaagcc	747
Db	62	TACGGCGCGCTCCCTCTTCGGGCTGTGATGGGTCACAGGCGATGGGTGTGGCCAAAGCGC	121
OY	748	ggtttggtctatttcaatggccggattatygctatgtcggaggaatgattaccttaacca	807
Db	122	GGCCCTCGTCTCTTCAACCAACCACTCTTTAGCCATGTCCGAAGATGACTTACCCCTACAT	181
OY	808	gttaagatcaactcccaatgtgaatttaaaacggttgytgcgttgattttagtagaca	867
Db	182	CTCCGCAATACCCCAAGGGGAGATTATACCAACCGTGGCGGTTACCACTTTAAGGGCCAG	241
OY	868	ttagaatcacacatgattgcgccaccggaagtcgacccggaatcgcgttgaaactcttcgct	927
Db	242	TTAAATTCACAAATGATGCGTCACACCCCAAACTGCACCCCGTGACCAACAACCTCCAGCGC	301
OY	928	ttaaagctacgagcgtcttcaaaagcccttaacctaaabaacttcggatttccaccgga	987
Db	302	CTCTCTCTACGACGTGTGCCAGAAACCATATCTCAAGTACTTCGGCTTCTCCACAGCGGC	361
OY	988	actaatatccggagcgtcgcgagattcagccttgaatcagcgaacgagtatgaacgatttcgg	104
Db	362	GTCAAAATCCCCCGACGTGCCAAATTCCTTTAAAGAAACCCACMTGATGACGATTTTCGGC	421
OY	1048	attcaagagaacttcgctcgttcgtaacctgcagcagaagtcgtttcaagctgcggagatg	1107
Db	422	ATCACCGAAGACTTCGTCTGTCTGCCGACACAGAGGTGGTCTTCAAGTTATCAGAATG	481
OY	1108	atccgcggttggtctccggtggtttacagacaagaacaaagtcgcaagattcggatttta	1167
Db	482	ATCACCGGAGGCTCCCGCTGTGTCTACGACAAAGATTTCAGATTCCGGAATTC	541
OY	1168	gacaacaatcgcggaat	1185
Db	542	GACACAGATGCTAAAGAT	559

```

RESULT 10
BE432853
LOCUS BE432853 520 bp mRNA linear EST 18-MAY-2001
DEFINITION E9399478 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone CLEB1167, mRNA sequence.
ACCESSION BE432853
VERSION BE432853.1 GI:9430792
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 520)
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
,S.D.
Generation of ESTs from tomato fruit tissue, breaker stage
JOURNAL Unpublished (2000)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source
1..520
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEB1167"
/clone_1lb="tomato breaker fruit, TIGR"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSMCadapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT 156 a 100 c 116 g 148 t
ORIGIN
Query Match 17.6%; Score 316.6; DB 10; Length 520;
Best Local Similarity 75.9%; Pred. No. 2.2e-83;
Matches 391; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 829 gattcaaacggttgctggttcgatttgatgacaaatagaaatcacatgatggc 888
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2 GATCTTAAACAGAGGGTGATTCGATTCGACGGCCACTAAATCCACATGATAGCT 61
QY 889 caccggaagtcgacccggaatcgcgtggaactcttcgcttaagctagcagctgttca 948
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62 CACCCAAAGCTGACCCAGCTTCGCGTAGCTATTGCTCTAGCTACGATGATTCAG 121
QY 949 aagccttactaaataacttcgattctacccggaagcacaataacacggaagtcgag 1008
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 122 AAGCCATACCTCAAGTACTTCAGATTTCATAAAAAATGGGAAAAATCAATATGTTGA 181
QY 1009 attcagctgacacgcaacagatgacagatcttcgagatcacagagaactctgcgc 1068
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 182 ATTCCAGTTGAAGACCAACATGATGATTTCCGCAATTACGAGAACTTCGCGTC 241
QY 1069 gtacctgacagcaagtcgttttcaagctgcggaagatgacgcggtgtctccggt 1128
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 242 ATTCCTGATCAACAAGTCGTTTCAAGATGCTGAATATGATCCGAGAGGTTACACG 301
QY 1129 gtttcgcaagaagaagtcgcaagatctcgagatttgagacaataagccgaagattca 1188
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 302 GTTTAGCAACGAACAAAGTTTCCGATTTCGGATTTCGGATAGTACGGAAGATGGG 361

```

```

QY 1189 tcgaacatgaatgcatgcatccacagatcttcctccatctcgaagccttg 1248
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 362 TCTGATTTGAATGGGTTCAAGTACTGATTTCTGTTCTGTTCCACTCTGGAATGCTTG 421
QY 1249 gaagagccagaacagatgaatgctgctgataaggtctctgatactccacagactca 1308
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 422 GAAGAAGCAGAAACAGATGAATTCGTTGTAATGGTTCATGATGACACCAACAGACTCC 481
QY 1309 atttcaacgagctcgaagagaatctcaagatgt 1343
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 482 ATTTCAATGAATGATGATGAAGGCTMAAGACTGT 516
RESULT 11
BH559796
LOCUS BH559796 742 bp DNA linear GSS 14-DEC-2001
DEFINITION BOHRR47TF BOHR Brassica oleracea genomic clone BOHRR47, DNA
sequence.
ACCESSION BH559796
VERSION BH559796.1 GI:17811576
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 742)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
unpublished (2001)
JOURNAL Other GSSs: BOHRR47TR
COMMENT Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
FEATURES
source
1..742
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOHRR47"
/clone_1lb="BOHR"
/note="Vector: pHOS1, site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 190 a 235 c 117 g 200 t
ORIGIN
Query Match 17.4%; Score 313.8; DB 12; Length 742;
Best Local Similarity 84.1%; Pred. No. 1.9e-82;
Matches 354; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 1 atggtcttcttaagggcaagcgtcgtgttctggaagatggtcgtggtgcaatactact 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 321 AAGGCTTTCTTACCGCGCATACGCGGCTTCTCGGAATGGGTGGTGAATATATACT 380
QY 61 cagcgccatatactgctctcttcaagcctcgactgagattatgtgactcttactctg 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 381 AAACCAACCTTATTCGTTCTTCTAAAGCTCCGCTTGCGTTATTCCTCCGTAACCAATG 440
QY 121 gccagctgtgtacacgaatgaatgcttcaatgttcatctgagcttcaacactccagctct 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 441 ACAAAATCGTCCCAAGAAAGCTCAATGTTCTCTGCGCTTCACACATCATCCGCTCTC 500
QY 181 catctccatgaacatcatcaaacctcccgccatctgttgttaagcccaagaagaa 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 501 CATTTCCCAACCAATCTCCACCTCTCCGCAATTGTTGTGAACCCAAACCAAGAA 560

```

Qy	241	tccaactaacaacgatgaatttgccttcagagagagcgcgcgcgctgttgaagcgcg	300
Db	561	ttccgcacacaaacagatgaactgtttccaaaaagcgcgcgcgctgttgacgcgcg	620
Qy	301	gagggcttcctgtcgaagcagcagaagatcacccgcgttcctaaagcgctgatactagt	360
Db	621	gaaggtttctctgtgtagcgcacgaggaagcagatccccctcccaaaacggcgatcttagc	680
Qy	361	gtccagatccgcggaaatttgcctcggatgatgaacgcccgcttcggcgatattccg	420
Db	681	gttcaaatccgcggaaacttgcctccgggtgaacgaacgacctgtccgttaaacctccc	740
Qy	421 g 421		
Db	741 g 741		
RESULT 12			
LOCUS	BH458011/c	594 bp	DNA linear GSS 12-DEC-2001
DEFINITION	BOHB102TF BOHB Brassica oleracea genomic clone BOHB102, DNA sequence.		
ACCESSION	BH458011		
VERSION	BH458011.1		
KEYWORDS	GSS.		
SOURCE	Brassica oleracea.		
ORGANISM	Brassica oleracea Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.		
REFERENCE	Town,C.D., Van Aken,S., Uffnerback,T. and Fraser,C.M. Whole genome shotgun sequencing of Brassica oleracea unpublished (2001)		
AUTHORS	Other-GSSs: BOHB102TR		
TITLE	Contact: Chris Town		
JOURNAL	TIGR		
COMMENT	9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523 Fax: 301-838-0208 Email: cdtown@tigr.org DNA is from a doubled haploid provided by Tom Osborn. Seq primer: TF Class: sheared ends.		
FEATURES	location/qualifiers		
SOURCE	1..594		
	/organism="Brassica oleracea"		
	/strain="T01000DH3"		
	/db_xref="taxon:3712"		
	/clone="BOHB102"		
	/clone.lib="BOHB"		
	/note="Vector: pHD1. Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHD1 using BstXI linkers"		
BASE COUNT	138 a 170 c 131 g 155 t		
ORIGIN			
Query Match	17.4%;	Score 312.8;	DB 12; Length 594;
Best Local Similarity	73.9%;	Freq. No. 3.3e-82;	
Matches 414; Conservative	0; Mismatches 137;	Indels 9; Gaps	1;
Qy	1047	gattcagagaaacttcgctcgtctcgtacccgagcagaagtcggttcctcaagctcgagagat	1106
Db	594	gattcagcggagaaacttcgctcgtctcgtccgacccagcaagatgcttccagctacggagat	535
Qy	1107	gattccgaggttgagctcctccggtgtttacgacaagaagaagtcgaagatcgggattt	1166
Db	534	gattccgaggttgagctcctccggtgtttacgacaagaagaagatcgaagatcgggattt	475
Qy	1167	agacaaataacgcgaagattcacaacataaagtggagttgattgctccaagattgcttctg	1226
Db	474	gagcgaataatgcgtgaagacgcttgcgttcgattcgcggtgagtcgagtgagagactgcttctg	415

QY	1227	cttcacatctctgggaacgcttgggaagagccagaacaacagatgaagtcgctcgatagatggtc	1286
Db	414	ttttccatctcttgggaacgcttgggaagagccagagacacaaacagagctgcgtcatcgggtc	355
QY	1287	ctgtatgactccacagaagactcaatttcaacagagctcagagaaatctcaagaagtgctc	1346
Db	354	atgcattgacgccgcccgacctgcatgtttacacacacacacgaacacatttcagagtgcttt	295
QY	1347	gtctgaatccgcctctgaatctcaaaaaccggtgaatcaactcgcgctccgcatctccaa	1406
Db	294	gncggagatfaaacactaaacctgaanaacagggagatccacgcgctcgccggttatctccga	235
QY	1407	cgaagatacaacagctcaacctcgaagcagggatgtgtcaacagaaacatgctcgcgcgttaa	1466
Db	234	gaaa-----ttaaattctgaagccggatggttaaacccgaatcttttagtgtaaaaa	184
QY	1467	aaccgaatccgcttaacttgacttagccgagcgcgctgagcctaagctcaagatctcgctaa	1526
Db	183	aacggcggttcgcttaccgtttgcttttaacccaacacggcgccttaagtgctcgggttcgcaa	124
QY	1527	agttgactcactactatggaaggttaagaacaatcttaccgcgatgaacgctttagcagag	1586
Db	123	actgacacttatctacacggagaggttcaaaaataatattatagctgacaggttaagttcgagag	64
QY	1587	agagcctcgtttcccg 1606	
Db	63	agagcctcgtttatgcccg 44	
RESULT	13		
LOCUS	B1974879	547 bp	mRNA linear EST 30-NOV-2001
DEFINITION	sa174db1.v1 Gm-cl068 glycine max cdna clone GENOME SYSTEMS CLONE		
ACCESSION	ID: Gm-cl068-4294 5' similar to tr:024023 024023 NEXXANTHIN		
VERSION	CLEAVAGE ENZYME: ; mRNA sequence.		
KEYWORDS	B1974879		
SOURCE	B1974879.1 GI:16349284		
ORGANISM	soybean.		
REFERENCE	Glycine max		
AUTHORS	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.		
1 (bases 1 to 547)			
Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Corryell,V., Khanna,A., Bolla,B., Marita,M., Hillier,L., Kuehba,T., Martin,D., Beck,C., Wylie,T., Underwood,K., Steptoe,B., Thelsging,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.			
Public Soybean EST Project			
Unpublished (1999)			
Contact: Shoemaker R/Public Soybean EST Project			
Public Soybean EST Project			
Washington University School of Medicine			
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA			
Tel: 314 286 1800			
Fax: 314 286 1810			
Email: est@watson.wustl.edu			
Possible reversed clone: similarity on wrong strand This clone is available through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway Huntville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccuteresgen.com			
High quality sequence stop: 423.			
Location/Qualifiers			
1..547			
/organism="Glycine max"			
/db_xref="taxon:3847"			
/clone="GENOME SYSTEMS CLONE ID: Gm-cl068-4294"			
/clone_lib="Gm-cl068"			
/tissue_type="Leaf, drought stressed, 1 month old plants,			
FEATURES			
SOURCE			

```

greenhouse grown"
/Lab_host="DH10B"
/note="vector: pluescript II SK+, Site_1: EcoRI, Site_2:
XhoI; The cDNA library was constructed from mRNA isolated
from drought stressed leaf tissue of the cultivar Williams
82. The month old greenhouse grown plants were deprived of
water for 3 days prior to harvesting the stressed leaf
tissue. Complementary DNA was synthesized from mRNA using
a primer consisting of a poly(dT) sequence with a XhoI
restriction site. EcoRI adaptors were ligated to the
blunt-ended cDNA fragments followed by XhoI digestion. The
cDNA fragments were directionally cloned into the
EcoRI-XhoI restriction site of the pluescript vector. The
ligated cDNA fragments were transformed into DH10B host
cells (GibcoBRL). This library was constructed in the
laboratory of Dr. Randy Shewmaker."

```

Query Match	17.28	Score 309.4	DB 10	Length 547
Best Local Similarly	74.48	Pred. No. 3.2e-81		
Matches 404; Conservative	0	Mismatches 136	Indels 3	Gaps 1

OY		991	aatcacccgagcgtctgagattcaagtcagtcatgcacgaagatgatccagattcgagatt	1050
Db		541	AAATCCCCCGAGCTGGAAATTCCCTTGAAGAAGACCACCATGATGCAGATTTCGCCATC	482
OY		1051	acagaacaattcgttgcgtctgataccagcaagaatcgtttccaagctcgcggagatgac	1110
Db		481	ACCGGAACCTTGCTGTGTGCTCCGCCACCAAGCGAGGTGCTTCATACTATACGAATATGATC	422
OY		1111	cgcggttgagctccggttggtttaagacaagaacaagatgcgaagtttgtgaatttaaac	1170
Db		421	ACCGGAGGCTCCCCCGTTGTCTACGACAAGAACAGATTTCCAGATTTGGGATTTCTGAC	362
OY		1171	aaatacgcgcgaagattcatacgaacattaagltgatalgtctcagattgcttcgtctc	1230
Db		361	AAGATTCGTAAAGATGCAAAATGATATGAAATGATGATGCTGCCGAGTGTCTGTTTC	302
OY		1231	catctctygaacgcgtttggygaagagccagaagaacagatgaatgcgtctgatataggctcgt	1290
Db		301	CACCTCTGGAAAGCGCTGGGAGAGCGCGGAATAATGATGATGATGCTGATCGTTCCCTGC	242
OY		1291	atgactccaccagcaatcaattttaaagagatctgaagatcccaagatgctcgtct	1350
Db		241	ATGACCCCCCGGAGCTCCATTTTCAACCAATGCGAAGAGATTTTAAGAAGCATCTTGTCG	182
OY		1351	gaaatccgcgtgaatctctaanaaccggttgatcaaatctgcctlcgatactcatccaagaa	1410
Db		181	GAGATTAAGGCTGAATTGTGAAGACAGGCAAGTCCACGAGAAAAACCATATATGC---GGAA	125
OY		1411	gatacaacaagtcaacctcgaaagcagggatggtlcaacagaaacatgcttgcgcgttaaac	1470
Db		124	TCCGACAACAGTCAACTTGGAAAGCCGGGATGTGTCAACAATAAACCTCGGGAGAAAAGCG	65
OY		1471	aaatcgcgtcttgcgtcttaagcagagccgfygccttaagatctcaagatctgcgttaagt	1530
Db		64	AAGTTCGGGTACTTAGCACACTTGGGAGCGCTGTCTTAAGTGTTCGGGTTTTGCCAAGGTT	5
OY		1531	gat	1533
Db		4	GAT	2
RESULT	14			
LOCUS	BE437072			
DEFINITION	EST408190 tomato breaker fruit, TIGR lycopersicon esculentum cDNA	547 bp	mRNA	linear EST 18-MAY-2001
ACCESSION	BE437072			
VERSION	BE437072.1			
KEYWORDS	EST.			

SOURCE	tomato.
ORGANISM	Lycopersicon esculentum
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
AUTHORS	1 (bases 1 to 547) Alcala,J., Vrehob,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Niemann,W., Fraser,C.M., Martin,G.B., Giovannoni,J.V. and Tanksley,S.D.
TITLE	Generation of ESTs from tomato fruit tissue, breaker stage
JOURNAL	unpublished (2000)
COMMENT	Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html 5 prime sequence.

FEATURES		Location/Qualifiers
source	1..547	
	/organism="Lycopersicon esculentum"	
	/cultivar="TA96"	
	/db_xref="taxon:4081"	
	/clone="CLEG35K8"	
	/clone_lib="tomato breaker fruit, TIGR"	
	/tissue_type="Pericarp"	
	/dev_string="breaker"	
	/lab_host="SOLR"	
	/note="Vector: plusescriptSkMCuadapt; Site_1: EcoRI; Site_2: XhoI; Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of the fruit). Fruit were cut in half and the seeds and the locules were discarded prior to freezing the pericarp."	
BASE COUNT	140 a 133 c 120 g 154 t	
ORIGIN		
Query Match	17.1%: Score 308.6; DB: 10; Length 547;	
Best Local Similarity	72.8%: Pred. No. 5.6e-81;	
Matches 398; Conservative	0; Mismatches 149; Indels 0; Gaps 0;	
OY	431 aactcccgatccatcaaaagagatgataatgagcaagagagcaaccactcaagcagc	490
DB	1 AATATCCCAATGTGTTTAAGGCGTTTACGTTTCGAAGGAGACGTAACTCTTTTGTAC	60
OY	491 cggtagcagtcacactcttcgcagcagagacggtatgctcaagccgctcaactcgaac	550
DB	61 CAACGCCGGACACCAATTTCCTTCGACGGCGACAGATGTGTTACGCCCTTAATTCAAA	120
OY	551 acggttcagctagctagccttcgcggttactcagactaacccggtttgttcaggaagtc	610
DB	121 ATGGGTGGGTAGTTACGCTTGCGCTTTCACGTGAAGACAGAGAGGCTGTGTAAGAAAG	180
OY	611 aattggtcgcagcgggtttccccaagacatcggtgagctcgaagccacacggtatg	670
DB	181 CTTTGGGTGCGCCTTTTTCCTTAAGGCATTTGGAATTCATGTGTCACCTCGGAATTG	240
OY	671 cccgactatctatcttcacccaagatgcagccggtatagtcgaaccgcgcacagga	730
DB	241 CAAGGCTTATGCTGTTTACGCTGCTGGGCTCTTCGGCACTGTGTGATCAGTAAAGAA	300
OY	731 ccggtgtagtaacgcggtttggtcctatccaatgagcgggttatgtgctatgctcgag	790
DB	301 CTGGGTGTCGAACGCCCGGTTTAGCTTATTTCATTAACCGATTACTTCCTTATGCTGAG	360
OY	791 atgattacttaacaaatctcagatcactcccaatgagatlttaaaacggtttgtcgt	850
DB	361 ATGATTGCTTACATGATAAAGGTAAACCCACCGCGGATCTTAAACACAGAGGTTGAT	420
OY	851 tcatatttgatgacaattagaaatcacatgatgtccaccgcgaagaatgcgaac	910
DB	421 TCGATTGTGACGGCCAGGTAAATCACCAGATGATGCTACCCCAAGCTCACCAGATT	480

QY	911	ccgggaactctcgcttcaagctcgcgcgcgtcttcaaacgcttactaaatcctcc	970
Db	481	CCGGGACACTATTTCCTCTTACTACGATGTCATTCAGAAAGCCATCTCAAGTACTTCA	540
QY	971	gattctc	977
Db	541	GATCTC	547
RESULT	15		
LOCUS	BE458861	617 bp	mRNA
DEFINITION	BE458861 tomato developing/immature green fruit Lycopersicon		
ACCESSION	BE458861	esculentum cDNA clone cLEM4F10, mRNA sequence.	
VERSION	BE458861.1	GI:9503163	
KEYWORDS	EST.		
ORGANISM	tomato.		
ORGANISM	Lycopersicon esculentum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;		
AUTHORS	1 (bases 1 to 617) Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.		
TITLE	Generation of ESTs from tomato fruit tissue, immature, green		
JOURNAL	unpublished (2000)		
COMMENT	Contact: CUGI Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html 5 prime sequence.		
FEATURES	location/Qualifiers		
SOURCE	1..617 /organism="Lycopersicon esculentum" /cultivar="TA96" /db_xref="taxon:4081" /clone="cLEM4F10" /clone_1fb="tomato developing/immature green fruit" /tissue_type="fruit" /dev_stage="immature green (5-35 days post-anthesis)" /lab_host="SOLR" /note="Vector: pBluescriptSKmCuaapt; Site_1: EcoRI; Site_2: XhoI; Fruit were tagged at 5 dpa (0.5 cm) and harvested at 7 day intervals through 35 dpa. Equal masses of tissue from each stage were combined (including seeds and locules) prior to mRNA isolation."		
BASE COUNT	164 a 145 c 137 g 171 t		
ORIGIN			
Query Match	16.9%: Score 303.4; DB 10; Length 617;		
Best Local Similarity	68.6%: Pred. No. 2.2e-79;		
Matches	418; Conservative 0; Mismatches 191; Indels 0; Gaps 0;		
QY	247	actaaacagatgaattgttccagagagcgcgcgcgacgcttgacgcgagcgagagt	306
Db	9	ACTTCACAGTGAATTTAGTCAGAAAGACGACGAGATGCGTTTAGATGCTGTAGAAAGT	68
QY	307	ttccctgtcgcacgaagagactaacaccgcgtctcttaaaagcgctgactctagtctcag	366
Db	69	GCTTTAACTAAACATGACACTTGACACCCCTTGCCGAAAGACCGACCCACGAGATCCAG	128
QY	367	atcgcgcggaattgtcctccgctgagatgaacagccgctccgcgcgtaactctccgctgtgc	426
Db	129	ATTTCGGGAATTTTGTCTCCGCTACCGGAAATTCACAGTCTGTCAATCTCTTCGGGTACC	188
QY	427	ggaaaaactcccgatcatcaaaagagbgtgltgtgcgaacgagacaaaccactctaac	486

Db	189	GGAAATAATACCCAATATGTTCTCAAGCGCTTACGTTGAAACGGAGCTAAOCCTTTT	248
QY	487	gagccgylgacaaagltcaacctcttcgcagagacgylatgltcacgcgylcaatl	546
Db	249	GAACCAACCGCGCGACACCATTTCTTGACGGGACGGATGTTCAACGCGTCAATTC	308
QY	547	gaacacgylttagctagctacgcttcgcgylttatcctaactaacgylttgttcagaa	606
Db	309	AAAAATGGGTGGGTAGTTAGTGGCTTCCGTTTCACTGAACAGAGAGCGTTGTCAAGAA	368
QY	607	cgtaattgggtgcacggytttcccaagaacatcgtgtagcttcaacgycacacggt	666
Db	369	AAAGCTTTGGGTCGCCCTGTTTCCCTTAAACCCATTTGGTGATTACATGGCACTCTTGA	428
QY	667	attgcgcacatcagltcatcttcaacgcgaagctgaagcggatagtcgaaccgcacac	726
Db	429	ATTGCAAGCGCTTATGCTTTTATACGCTCGTGGGCTCTTGGACTGTTGATCAACAGTAA	488
QY	727	ggaaccgyltagtgcacaaacgcgylttgltctatltcaatgycgcgyltatgltacgtc	786
Db	489	GGAACTGGTGTGGCAAAAGCGCGGTTTATCTATTTCATATAACGATTACTTGATATGCT	548
QY	787	gagagatgattacctaacttacaagltcagaatcattcccaatggagaattaaaaacgyltgt	846
Db	549	GAAGATGATTTTGCCTTACCATGTAAAGGTATACACCCACGCGGATCTTAAAAACGAAAGT	608
QY	847	cggttcgat 855	
Db	609	CGATTTCGAT 617	

Search completed: July 24, 2002, 06:53:47
Job time: 5972 sec

.....

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 03:54:00 ; Search time 57.66 Seconds
(without alignments)
7668.059 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800

Sequence: 1 atgcgtcttcacgcgaac.....tgcggaagcagtcgtgtga 1800

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*

2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*

3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*

4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*

5: /cgn2_6/ptodata/1/ina/PCrUS.COMB.seq:*

6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38.8	2.2	7218	1 US-08-232-463-14	Sequence 14, Appli
2	34.8	1.9	4695	6 5225348-3	Patent No. 5225348
3	34.2	1.9	1221	1 US-08-434-881-1	Sequence 1, Appli
4	34.2	1.9	1221	3 US-08-977-771-1	Sequence 1, Appli
5	34.2	1.9	1221	4 US-09-361-773-1	Sequence 1, Appli
6	34.2	1.9	1370	4 US-08-026-408-12	Sequence 12, Appli
7	34.2	1.9	1371	4 US-09-026-408-1	Sequence 1, Appli
8	33.4	1.9	788	4 US-08-991-789A-177	Sequence 177, App
9	33.4	1.9	788	4 US-09-062-451-177	Sequence 177, App
10	33.4	1.9	1047	2 US-08-494-907-11	Sequence 11, Appli
11	33.4	1.9	1047	5 PCT-US96-10986-11	Sequence 11, Appli
12	33.4	1.9	1753	6 5225348-2	Patent No. 5225348
13	33.4	1.9	3680	2 US-08-494-907-1	Sequence 1, Appli
14	33.4	1.9	3680	5 PCT-US96-10986-1	Sequence 1, Appli
15	33.4	1.9	5076	2 US-08-494-907-2	Sequence 2, Appli
16	33.4	1.9	5076	5 PCT-US96-10986-2	Sequence 2, Appli
17	33.4	1.9	6170	2 US-08-494-907-4	Sequence 4, Appli
18	33.4	1.9	6170	5 PCT-US96-10986-4	Sequence 4, Appli
19	33.4	1.9	6387	2 US-08-494-907-3	Sequence 3, Appli
20	33.4	1.9	6387	5 PCT-US96-10986-3	Sequence 3, Appli
21	33.4	1.9	7198	4 US-08-994-035C-4	Sequence 4, Appli
22	33.4	1.9	4411529	4 US-09-103-840A-1	Sequence 1, Appli
23	33	1.8	607	4 US-09-385-982-525	Sequence 525, App
24	33	1.8	2128	2 US-08-371-377-15	Sequence 16, Appli
25	33	1.8	80161	3 US-09-036-987A-1	Sequence 1, Appli
26	33	1.8	80161	4 US-09-370-700-1	Sequence 1, Appli
27	32.6	1.8	1092	4 US-09-077-675A-15	Sequence 15, Appli

C	28	32.6	1.8	3129	4	US-09-077-675A-14	Sequence 14, Appl
	29	31.8	1.8	7218	1	US-08-232-463-14	Sequence 14, Appl
	30	31.6	1.8	1876	3	US-08-714-918-33	Sequence 33, Appl
	31	31.6	1.8	1876	4	US-09-265-315-33	Sequence 33, Appl
	32	31.6	1.8	1876	4	US-09-265-315-33	Sequence 33, Appl
	33	31.6	1.8	1876	4	US-09-266-417-33	Sequence 33, Appl
C	34	31.4	1.7	30001	2	US-08-125-468-1	Sequence 1, Appl
C	35	31.4	1.7	30001	2	US-08-474-933-1	Sequence 1, Appl
C	36	31.4	1.7	3048	2	US-08-680-326-25	Sequence 25, Appl
C	37	31.2	1.7	3077	1	US-08-074-121-1	Sequence 1, Appl
C	38	31.2	1.7	3077	5	PCT-US94-06447-1	Sequence 1, Appl
C	39	31	1.7	289	4	US-09-007-005-17	Sequence 17, Appl
C	40	31	1.7	289	4	US-09-244-796-17	Sequence 17, Appl
C	41	30.4	1.7	551	4	US-09-328-111-448	Sequence 448, Appl
C	42	30.4	1.7	4316	1	US-08-317-450B-14	Sequence 14, Appl
C	43	30.4	1.7	4316	3	US-08-800-593-14	Sequence 14, Appl
C	44	30.4	1.7	5200	1	US-08-317-450B-12	Sequence 12, Appl
C	45	30.4	1.7	5200	3	US-08-800-593-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHIEFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOULPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
CLASSIFICATION: 435
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
SEQUENCE CHARACTERISTICS:
SEQUENCE FOR SEQ ID NO: 14:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgpt-Fls
US-08-232-463-14

```

Query Match      2.2%; Score 38.8; DB 1; Length 7218;
Best Local Similarity 7.9%; Pred. No. 0.049;
Matches 16; Conservative 112; Mismatches 74; Indels 0; Gaps 0;

QY 21 gctgcggttcctcgagatgctgtgcaatcatctacgcgcgcatctgcttc 80
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 1041 GCGTCGACAGTCGAGGAGCTGCGATYYYYYYYYYYYYYYYYYYYYYY 1100

QY 81 tcaagcgcgatgagttattgtgctccttacctatgctgagcagtcgtgacacgttaa 140
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1101 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1160

QY 141 gctcaatgttcatctgcgtcgtcacactccctccagctctcattccctaaagcaatc 200
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1161 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1220

QY 201 aaactcccgccatgtgtc 222
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1221 YYYYYYYYYYYYYYYYYYYY 1242

RESULT 2
5225348-3
; Patent No. 5225348
; APPLICANT: HAGATA, SHIGEKAZU; SUGANO, SUMIO; KIM, DONG W.;
; UETSUKI, TAICHI; KAZIRO, YOSHITO
; TITLE OF INVENTION: DNA FRAGMENT AND EXPRESSION PLASMID
; CONTAINING THE DNA FRAGMENT
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/447,823
; FILING DATE: 08-DEC-1989
; SEQ ID NO: 3:
; LENGTH: 4695
5225348-3

Query Match      1.9%; Score 34.8; DB 6; Length 4695;
Best Local Similarity 48.1%; Pred. No. 0.78;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 340 cctaaacggtgactagttgtcagatcgcgcggaattgtgctcggtagaataacag 399
    || | | | | | | | | | | | | | | | | | | | | | | | |
Db 3386 cctgattgattgcacacgctcaccatgcatgcaagttgctgagctgaagaaag 3445

QY 400 cccgcgcgggaatctccggtgtgctggaactcccgatccatcaagagtgat 459
    | | | | | | | | | | | | | | | | | | | | | | | |
Db 3446 attgacgcgctctgtaaaaaagctggaagatgcccctaaattcttgaagtgtgat 3505

QY 460 gtgcgaacgagctaacccacttcaacgagcgcgtagacagtcacacacttccagcaga 519
    | | | | | | | | | | | | | | | | | | | | | | | |
Db 3506 gctgccattgtgatatagttcttcgcaagcccatggtgtgtagagacttccagactat 3565

QY 520 gacggatggtacgcgcgtcaaat 545
    | | | | | | | | | | | | | | | | | | | | | | | |
Db 3566 ccactttgggaagatgactactt 3591

RESULT 3
US-08-434-881-1/c
; Sequence 1, Application US/08434881
; Patent No. 5804376
; GENERAL INFORMATION:
; APPLICANT: Braxton, Scott M.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Pancreas-Derived Serpin
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3330 Hillview Avenue
; CITY: Palo Alto
; STATE: California

```

```

; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,881
; FILING DATE: Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF0035 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1221 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; LIBRARY: Pancreas
; CLONE: 222689
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1221
; US-08-434-881-1

Query Match      1.9%; Score 34.2; DB 1; Length 1221;
Best Local Similarity 54.3%; Pred. No. 0.51;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 97 agttgtgagtccttacctatgagccagtcgtgcacagtaagctcaatgttcaatct 156
    | | | | | | | | | | | | | | | | | | | | | | | |
Db 788 ACTTCTCTATATCCATACCTTCTGACGAAAGTATGATATTAAGCTAATTCATCACCCT 729

QY 157 gcgcctacactccctccagctcttcatctccctaaagcaatcaaacctccgcgcat 216
    | | | | | | | | | | | | | | | | | | | | | | | |
Db 728 TTGTAAGACAATTCTAAACTGTGATGTTTCAGGGAAGATTGAAATAACCATATTTT 669

QY 217 gttgtta 223
    | | | | |
Db 668 GTTCTCA 662

RESULT 4
US-08-977-771-1/c
; Sequence 1, Application US/08977771
; Patent No. 6013448
; GENERAL INFORMATION:
; APPLICANT: Braxton, Scott M.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Pancreas-Derived Serpin
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3330 Hillview Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30

```

```

:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/977,771
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/434,881
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Luther, Barbara J.
: REGISTRATION NUMBER: 33954
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-852-0195
: TELEFAX: 415-852-0195
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1221 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: IMMEDIATE SOURCE:
: LIBRARY: Pancreas
: CLONE: 222689
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1221
:
: US-08-977-771-1

```

```

Query Match      1.9%; Score 34.2; DB 3; Length 1221;
Best Local Similarity 54.3%; Pred. No. 0.51;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY  97 agtatttgagctccttaccatgagcagtcgtgcacacgtaagctcaatgttcatct 156
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  788 ACTCTTCTATATCATCTCTCTCTGCGAGAGTATGATTAATTAACCTAATTCATCACCCT 729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  157 ggccttcacacccctccacgtcttcatctccctaaagcaatcacaactcccgccatt 216
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  728 TTGTAAGCAATTTCAAAACTTGGTAGTTTCAGGAGATTCAGAAAAATTAACCATATTTT 669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  217 gttgtta 223
    |||||
DB  668 GTTCTCA 662

```

```

RESULT 5
US-09-361-773-1/c
: Sequence 1, Application US/09361773
: Patent No. 6197519
: GENERAL INFORMATION:
: APPLICANT: Braxton, Scott M.
: APPLICANT: Wilde, Craig G.
: APPLICANT: Diep, Dinh
: TITLE OF INVENTION: Pancreas-Derived Serpin
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3330 Hillview Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PC-DOS/MS-DOS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/361,773
: FILING DATE:
: CLASSIFICATION:

```

```

:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/977,771
: FILING DATE:
: APPLICATION NUMBER: 08/434,881
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Luther, Barbara J.
: REGISTRATION NUMBER: 33954
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-852-0195
: TELEFAX: 415-852-0195
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1221 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: IMMEDIATE SOURCE:
: LIBRARY: Pancreas
: CLONE: 222689
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1221
:
: US-09-361-773-1

```

```

Query Match      1.9%; Score 34.2; DB 4; Length 1221;
Best Local Similarity 54.3%; Pred. No. 0.51;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY  97 agtatttgagctccttaccatgagcagtcgtgcacacgtaagctcaatgttcatct 156
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  788 ACTCTTCTATATCATCTCTCTCTGCGAGAGTATGATTAATTAACCTAATTCATCACCCT 729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  157 ggccttcacacccctccacgtcttcatctccctaaagcaatcacaactcccgccatt 216
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  728 TTGTAAGCAATTTCAAAACTTGGTAGTTTCAGGAGATTCAGAAAAATTAACCATATTTT 669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  217 gttgtta 223
    |||||
DB  668 GTTCTCA 662

```

```

RESULT 6
US-09-026-408-12/c
: Sequence 12, Application US/09026408
: Patent No. 6303338
: GENERAL INFORMATION:
: APPLICANT: Ni et al.
: TITLE OF INVENTION: PANCREAS-DERIVED PLASMINOGEN ACTIVATOR
: TITLE OF INVENTION: INHIBITOR
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 NEW YORK AVENUE, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: USA
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/026,408
: FILING DATE: Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/934,011
: FILING DATE: 15-AUG-1997

```

```

PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/024,056
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0300002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1370 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 67..1281
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 67..120
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..1281
US-09-026-408-12

```

Query Match	1.9%	Score 34.2;	DB 4;	Length 1370;
Best Local Similarity	54.3%;	Pred. NO. 0.55;		
Matches 69;	Conservative	0;	Mismatches 58;	Indels 0;
				Gaps 0;

Qy	97	agattatgtacactcttcaacctcagtcggaagtcgctgcacacogtaagaaccatgltcatct	156
Dd	851	ACTCTTCTATATCCATACCTTTCTGCAGGAAGTATGATTAACTTAATTCATACCT	792
Qy	157	gcgcttacactctccagcctttcatcttcccctaagaatacalcaaacctctccgccat	216
Dd	791	TGTGAAGCAATTTCTAANAACCTTGSTAGTTCAGGAGAATTCAGAAAATATACCATATTTT	732
Qy	217	gtcgta	223
Dd	731	GTTCTCA	725

RESULT 7
US-09-026-408-1/c
; Sequence 1, Application US/09026408
; Patent No. 6303338

GENERAL INFORMATION:
APPLICANT: NI et al.
TITLE OF INVENTION: PANCREAS-DERIVED PLASMINOGEN ACTIVATOR
TITLE OF INVENTION: INHIBITOR
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/026,408
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/934,011

```

FLING DATE: 15-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,056
FLING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0300002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1371 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 67..1242
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 109..1242
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 67..108
;
;
US-09-026-408-1

```

Query Match	1.9%	Score 34.2;	DB 4;	Length 1371;
Best Local Similarity	54.3%	Pred. No. 0.55;		
Matches 69; Conservative	0;	Mismatches 58;	Indels 0;	Gaps 0;

Qy	97	agatattgtaaccccttaccatacgtggccaagtcgcgtctcaagaagtgaactcaattgtttatc	156
Dd	851	ACTCTTCTATATCATCAACCTCTCGCAGAGAAGTATGTAATTAAAGCTAAATTTCATCACCT	7922
Qy	157	gcgcctcacacctcctccaaagctctctcatcttcctctaagaatcaatcaaacctctcccgcacat	216
Dd	791	TGTGAAGACAATTCTAATAAACCTTGTAATTCCAGGGAGAAATTCAGAAAAATTAACCATATTTT	7322
Qy	217	gttgta 223	
Dd	731	GTTCTCA 725	

RESULT 8
US-08-991-789A-177
; Sequence 177, Application US/08991789A

APPLICANT: Frudakis, Tony N.
Smith, John M.
Reed, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 292

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997

```

; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Potter, Jane E. R.
;   REGISTRATION NUMBER: 33,332
;   REFERENCE/DOCKET NUMBER: 210121.419C3
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 682-4900
;   TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 177:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 788 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 177:
US-08-991-789A-177

Query Match      1.9%; Score 33.4; DB 4; Length 788;
Best Local Similarity 48.2%; Pred. No. 0.71;
Matches 94; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 340 cctaaacggtcgtatccttagtgcagatcgccggaatttgcgcggtgaatgaacag 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 536 CCTGATTGGATTGCCACACGGCTCACAATTGCATGCAGTTGCTGAGAGGAGGAGG 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 400 cccgctcgagctaatcttcggtgctcggaacttcccgatccatcaagaagtgat 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 596 ATTGATGCCCTTCTGTGTAAGAAAGCTGGAAGATGGCCCTTAATTCTTAAAGTCTG 655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 gtgcgaacgagctaacccacttcacgagccggtgacaggtcacaccacttctcgagga 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 656 GCTGCCATTGTTGATATGTTCTGTGGCAAGCCCATGTGTGAGAGCTTCTCAGACTAT 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 520 gacggtatggttcac 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 716 CCACCTTTGGGTGCG 730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-09-062-451-177
; Sequence 177, Application US/09062451
; Patent No. 6344550
; GENERAL INFORMATION:
;   APPLICANT: Frudakis, Tony N.
;   APPLICANT: Smith, John M.
;   APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 297
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/062,451
; FILING DATE: 04-Apr-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Makl, David J.
;   REGISTRATION NUMBER: 31,392
;   REFERENCE/DOCKET NUMBER: 210121.419C2
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 622-4900
;   TELEFAX: (206) 682-6031
```

```

; INFORMATION FOR SEQ ID NO: 177:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 788 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-09-062-451-177

Query Match      1.9%; Score 33.4; DB 4; Length 788;
Best Local Similarity 48.2%; Pred. No. 0.71;
Matches 94; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 340 cctaaacggtcgtatccttagtgcagatcgccggaatttgcgcggtgaatgaacag 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 536 CCTGATTGGATTGCCACACGGCTCACAATTGCATGCAGTTGCTGAGAGGAGGAGG 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 400 cccgctcgagctaatcttcggtgctcggaacttcccgatccatcaagaagtgat 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 596 ATTGATGCCCTTCTGTGTAAGAAAGCTGGAAGATGGCCCTTAATTCTTAAAGTCTG 655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 gtgcgaacgagctaacccacttcacgagccggtgacaggtcacaccacttctcgagga 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 656 GCTGCCATTGTTGATATGTTCTGTGGCAAGCCCATGTGTGAGAGCTTCTCAGACTAT 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 520 gacggtatggttcac 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 716 CCACCTTTGGGTGCG 730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-08-494-907-11/c
; Sequence 11, Application US/08494907
; Patent No. 5955298
; GENERAL INFORMATION:
;   APPLICANT: Thomasow, Linda S
;   APPLICANT: Banger, Mahalaxmi
;   APPLICANT: Weller, David M
;   APPLICANT: Cook, R. James
; TITLE OF INVENTION: Sequences for Production of
; TITLE OF INVENTION: 2,4-Diacetylphloroglucinol and Methods
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Margaret A. Connor, USDA-ARS
; STREET: 800 Buchanan Street
; CITY: Albany
; STATE: CA
; COUNTRY: USA
; ZIP: 94710
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494,907
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Connor, Margaret A
;   REGISTRATION NUMBER: 30043
;   REFERENCE/DOCKET NUMBER: 0009,95
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (510) 559-6067
;   TELEFAX: (510) 559-5777
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1047 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
```



```

APPLICANT: Weller, David M
APPLICANT: Cook, R. James
TITLE OF INVENTION: Sequences for Production of
TITLE OF INVENTION: 2,4-Diacetylphloroglucinol and Methods
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Margaret A. Connor, USDA-ARS
STREET: 800 Buchanan Street
CITY: Albany
STATE: CA
COUNTRY: USA
ZIP: 94710
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/94,907
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Connor, Margaret A
REGISTRATION NUMBER: 30043
REFERENCE/DOCKET NUMBER: 0009.95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 559-6067
TELEFAX: (510) 559-5777
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas fluorescens
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (2889..3680)
OTHER INFORMATION: /note="ph1a", transcribed from
OTHER INFORMATION: right to left"
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (1210..1917)
OTHER INFORMATION: /note="ph1b", transcribed from
OTHER INFORMATION: right to left"
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (1689..2855)
OTHER INFORMATION: /note="ph1c", transcribed from
OTHER INFORMATION: right to left"
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (2..1048)
OTHER INFORMATION: /note="ph1d", transcribed from
OTHER INFORMATION: right to left"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2118..3371
OTHER INFORMATION: /note="to right"
OTHER INFORMATION: to right"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..3680
OTHER INFORMATION: /note="Seq ID NO:1 contains genes
OTHER INFORMATION: necessary for Ph1 synthesis."
US-08-494-907-1

```

[illegible]

```

RESULT 14
PCT-US96-10986-1/c
Sequence 1, Application PC/TUS9610986
GENERAL INFORMATION:
TITLE OF INVENTION: Sequences for Production of
TITLE OF INVENTION: 2,4-Diacetylphloroglucinol and Methods
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Stephan A. Pendorf, DOMINIK & STEIN
STREET: 600 N. West Shore Boulevard, Suite 1000
CITY: Tampa
STATE: FL
COUNTRY: USA
ZIP: 33609
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10986
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Pendorf, Stephan A.
REGISTRATION NUMBER: 32665
REFERENCE/DOCKET NUMBER: A700.320
TELECOMMUNICATION INFORMATION:
TELEPHONE: (813) 289-2966
TELEFAX: (813) 289-2966
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: pseudomonas fluorescens
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (2889..3680)
OTHER INFORMATION: /note="ph1b, transcribed from
OTHER INFORMATION: right to left"
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (1210..1917)
OTHER INFORMATION: /note="ph1b, transcribed from
OTHER INFORMATION: right to left"
FEATURE:
NAME/KEY: misc_feature

```


Search completed: July 24, 2002, 06:21:27
Job time: 8847 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2002, 10:11:48 ; Search time 34.28 Seconds

(without alignments)
3022.869 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTATAVNSGRWLGSGNHT.....VPYGHGTFIGADDLAKQV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.protist:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3150	100.0	599	10	Q9LR77 arabidopsis
2	3143	99.8	599	10	Q93ZU5 arabidopsis
3	2289	72.7	604	10	Q9M3Z9 solanum tub
4	2280.5	72.4	605	10	Q240Z3 lycopersico
5	2170.5	68.9	657	10	Q9M9F5 arabidopsis
6	2168.5	68.8	612	10	Q9F5Z4 arabidopsis
7	2165	68.7	615	10	Q9M6E8 arabidopsis
8	2066.5	65.6	625	10	Q9M6E8 arabidopsis
9	2044.5	64.9	589	10	Q9C6Z1 arabidopsis
10	1991	63.2	583	10	Q9A505 arabidopsis
11	1977	62.8	569	10	Q9AXZ3 persea amer
12	1930	61.3	604	10	Q24592 zea mays (m
13	1663.5	52.8	577	10	Q9LRM7 arabidopsis
14	968	30.7	595	10	Q49675 arabidopsis
15	959.5	30.5	543	10	Q94IR2 phaseolus v
16	938	29.8	538	10	Q9LY63 arabidopsis

17	937	29.7	538	10	Q65572 arabidopsis
18	837.5	26.6	200	10	Q94EN8 lactuca sat
19	798.5	25.3	524	10	Q9AXZ5 persea amer
20	745	23.7	446	10	Q49895 malus domes
21	528.5	16.8	483	16	Q9AA32 caulobacter
22	527.5	16.7	456	2	Q9AM11 streptomyce
23	526.5	16.7	456	2	Q9JFA4 streptomyce
24	466	14.8	501	16	Q06785 mycobacteri
25	464	14.7	503	2	Q9RK46 streptomyce
26	454.5	14.4	502	16	Q05905 mycobacteri
27	394	12.5	490	16	P74334 synechocyst
28	352	11.2	485	2	Q53353 pseudomonas
29	317	10.1	480	16	P74370 synechocyst
30	303.5	9.6	490	2	Q52008 pseudomonas
31	302.5	9.6	552	10	Q93305 oryza sativ
32	270	8.6	542	2	Q68868 synechococc
33	260	8.3	526	13	Q91993 gallus gall
34	256	8.1	516	13	Q90WH4 brachydanio
35	253	8.0	616	10	Q9M079 arabidopsis
36	244	7.7	618	10	Q9SHD9 arabidopsis
37	238.5	7.6	549	13	Q90WH3 brachydanio
38	238.5	7.6	556	5	Q9TX79 caenorhabdi
39	232.5	7.4	532	11	Q99NF1 mus musculu
40	231	7.3	483	16	Q9AB98 caulobacter
41	225	7.1	547	4	Q9HAY6 homo sapien
42	224	7.1	566	11	Q9JIS6 mus musculu
43	223	7.1	547	4	Q9NVH5 mus musculu
44	221	7.0	566	11	Q9ERN9 mus musculu
45	216	6.9	533	13	Q9Y125 ambystoma t

ALIGNMENTS

RESULT	ID	Q9LR77	PRELIMINARY:	PRT:	599 AA.
1	Q9LR77	01-OCT-2000 (TRENBLrel. 15, Created)			
AC	Q9LR77	01-DEC-2001 (TRENBLrel. 19, Last annotation update)			
DT	01-OCT-2000 (TRENBLrel. 15, Created)				
DT	01-DEC-2001 (TRENBLrel. 19, Last annotation update)				
DE	9-CIS-EPOXYCAROTENOID DIOXYGENASE (NEOXANTHIN CLEAVAGE ENZYME).				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:				
OC	Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:				
OC	eurosid II: Brassicales: Brassicaceae: Arabidopsis.				
OX	NCBI_Taxid=3702;				
RN	[1]	SEQUENCE FROM N.A.			
RP	RC	STRAIN-COLUMBIA:			
RC	RC	Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;			
RA	Submitted (Jun-1999) to the EMBL/GenBank/DBJ databases.				
RL	[2]	SEQUENCE FROM N.A.			
RP	RC	STRAIN-COLUMBIA:			
RC	RC	MEDLINE=20277480; PubMed=10819329;			
RX	RC	Medline=20277480; PubMed=10819329;			
RA	Nakamura Y.;				
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence				
RT	features of the regions of 4,504,864 bp covered by sixty p1 and TAC				
RT	clones.";				
RL	DNA Res. 7:131-135(2000).				
RN	[3]	SEQUENCE FROM N.A.			
RP	RC	STRAIN-COL.			
RC	RC	Iuchi S., Kobayashi M., Shinozaki K.;			
RA	"Characterization of neoxanthin cleavage enzyme from Arabidopsis				
RT	thaliana.";				
RT	Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.				
RL	EMBL: AB026817; BAB01336.1; -;				
DR	EMBL: AB026549; BAB70609.1; -;				
DR	EMBL: IP004294; RPE65.				
DR	PIfam: PF03055; RPE65; 1.				
KW	Dioxygenase.				

SQ SEQUENCE 599 AA: 65856 MW: 7D513E39945E0CF3 CRC64:

Query Match 100.0%; Score 3150; DB 10; Length 599;
Best Local Similarity 100.0%; Pred. No. 1.7e-242;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASFTATAAVSGRWLGNGHTQPPPLSSQSSDLSYCSLPMASHVTRKLVNSALHTPPAL 60
DB 1 MASFTATAAVSGRWLGNGHTQPPPLSSQSSDLSYCSLPMASHVTRKLVNSALHTPPAL 60
QY 61 HEPKSSNSPAIYVKKRAKESNTKQNNLQRAAAALDAAEGFLVSHKILHPLPKTADPS 120
DB 61 HEPKSSNSPAIYVKKRAKESNTKQNNLQRAAAALDAAEGFLVSHKILHPLPKTADPS 120
QY 121 VOIAGNFAVNEQPVRRNLPPVVGKLPDSIKGVYVRNGANPLHEPVYGNHFFPDGGMVHAV 180
DB 121 VOIAGNFAVNEQPVRRNLPPVVGKLPDSIKGVYVRNGANPLHEPVYGNHFFPDGGMVHAV 180
QY 181 KEHGSASVACREPTQTNRRFVQERQLGRPVFPKAIIGELHGHGTIARLMLFYARAAAGIVDP 240
DB 181 KEHGSASVACREPTQTNRRFVQERQLGRPVFPKAIIGELHGHGTIARLMLFYARAAAGIVDP 240
QY 241 AHGTGVAANGIYFNGRLLAISEDLPYQVQITPNNGDLKTVGFRFDDGQLESTMIAPKV 300
DB 241 AHGTGVAANGIYFNGRLLAISEDLPYQVQITPNNGDLKTVGFRFDDGQLESTMIAPKV 300
QY 301 DPESGELFALSYDVVSKPYLKTYFRSPDGTSPDVEIQDQPTMMDFAITENFVVVPDQ 360
DB 301 DPESGELFALSYDVVSKPYLKTYFRSPDGTSPDVEIQDQPTMMDFAITENFVVVPDQ 360
QY 361 QVYFKLPEMIRGSGPVVYVKNKVARFGIILDKYAEDSSNKKWIDAPDCCFCHLNAMEEPE 420
DB 361 QVYFKLPEMIRGSGPVVYVKNKVARFGIILDKYAEDSSNKKWIDAPDCCFCHLNAMEEPE 420
QY 421 TDEVVVIGSCMPDPSIFNESPENLKSYLEIRLNKKTGSESTRPPIISNEDQVNLKAG 480
DB 421 TDEVVVIGSCMPDPSIFNESPENLKSYLEIRLNKKTGSESTRPPIISNEDQVNLKAG 480
QY 481 VNRNMLGRKTKFAYLALAPWPKVSGFAKYVDLTTGEVKKHLXGDNRYGGEPLFLPGEGE 540
DB 481 VNRNMLGRKTKFAYLALAPWPKVSGFAKYVDLTTGEVKKHLXGDNRYGGEPLFLPGEGE 540
QY 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEAVTKLPSRVPYGFHGTFTGADDLAKQV 599
DB 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEAVTKLPSRVPYGFHGTFTGADDLAKQV 599

RESULT 2
Q93205 PRELIMINARY; PRT: 599 AA.

AC 093205;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE PUTATIVE 9-CIS-EPOXYCAROTENOID DIOXYGENASE.
GN MOA2.4/AT3G14440.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
RN NCB1
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
R Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA Ishida J., Jones T., Kaniya A., Karlin-Neumann G., Kawai J., Kim C.,
Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shim P.,
RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full Length cDNA of gene MOA2.4/AT3g14440 (GI:11994214).";

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY056255; AAL07104.1; -
KW Dioxigenase.
SQ SEQUENCE 599 AA: 65813 MW: B9007A2DC1C15506 CRC64:

Query Match 99.8%; Score 3143; DB 10; Length 599;
Best Local Similarity 99.8%; Pred. No. 6.3e-242;
Matches 598; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MASFTATAAVSGRWLGNGHTQPPPLSSQSSDLSYCSLPMASHVTRKLVNSALHTPPAL 60
DB 1 MASFTATAAVSGRWLGNGHTQPPPLSSQSSDLSYCSLPMASHVTRKLVNSALHTPPAL 60
QY 61 HEPKSSNSPAIYVKKRAKESNTKQNNLQRAAAALDAAEGFLVSHKILHPLPKTADPS 120
DB 61 HEPKSSNSPAIYVKKRAKESNTKQNNLQRAAAALDAAEGFLVSHKILHPLPKTADPS 120
QY 121 VOIAGNFAVNEQPVRRNLPPVVGKLPDSIKGVYVRNGANPLHEPVYGNHFFPDGGMVHAV 180
DB 121 VOIAGNFAVNEQPVRRNLPPVVGKLPDSIKGVYVRNGANPLHEPVYGNHFFPDGGMVHAV 180
QY 181 KEHGSASVACREPTQTNRRFVQERQLGRPVFPKAIIGELHGHGTIARLMLFYARAAAGIVDP 240
DB 181 KEHGSASVACREPTQTNRRFVQERQLGRPVFPKAIIGELHGHGTIARLMLFYARAAAGIVDP 240
QY 241 AHGTGVAANGIYFNGRLLAISEDLPYQVQITPNNGDLKTVGFRFDDGQLESTMIAPKV 300
DB 241 AHGTGVAANGIYFNGRLLAISEDLPYQVQITPNNGDLKTVGFRFDDGQLESTMIAPKV 300
QY 301 DPESGELFALSYDVVSKPYLKTYFRSPDGTSPDVEIQDQPTMMDFAITENFVVVPDQ 360
DB 301 DPESGELFALSYDVVSKPYLKTYFRSPDGTSPDVEIQDQPTMMDFAITENFVVVPDQ 360
QY 361 QVYFKLPEMIRGSGPVVYVKNKVARFGIILDKYAEDSSNKKWIDAPDCCFCHLNAMEEPE 420
DB 361 QVYFKLPEMIRGSGPVVYVKNKVARFGIILDKYAEDSSNKKWIDAPDCCFCHLNAMEEPE 420
QY 421 TDEVVVIGSCMPDPSIFNESPENLKSYLEIRLNKKTGSESTRPPIISNEDQVNLKAG 480
DB 421 TDEVVVIGSCMPDPSIFNESPENLKSYLEIRLNKKTGSESTRPPIISNEDQVNLKAG 480
QY 481 VNRNMLGRKTKFAYLALAPWPKVSGFAKYVDLTTGEVKKHLXGDNRYGGEPLFLPGEGE 540
DB 481 VNRNMLGRKTKFAYLALAPWPKVSGFAKYVDLTTGEVKKHLXGDNRYGGEPLFLPGEGE 540
QY 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEAVTKLPSRVPYGFHGTFTGADDLAKQV 599
DB 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEAVTKLPSRVPYGFHGTFTGADDLAKQV 599

RESULT 3
Q9M329 PRELIMINARY; PRT: 604 AA.

AC 09M329;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DE PUTATIVE 9-CIS-EPOXYCAROTENOID DIOXYGENASE.
GN NCEB1.
OS Solanum tuberosum (potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasteride I; Solanales; Solanaceae; Solanum.
NCBI_TaxID=4113;
RN NCB1
RP SEQUENCE FROM N.A.
RA Burdige A., Taylor I.B., Thompson A.;
RT "Potato putative 9-cis-epoxycarotenoid dioxygenase 1 cDNA.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ276244; CAB76920.1;
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.

RA [1]
RP SEQUENCE FROM N.A.
RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
RA Shun P., Altafi H., Bel B., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F3p9 from chromosome
RT 1";
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chlou J., Choi E.,
RA Conn L., Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC013430; AAF11797.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING: UNKNOWN_1.
SQ SEQUENCE 657 AA; 73015 MW; F41DECBE94806318 CRC64;

Query Match 68.9%; Score 2170.5; DB 10; Length 657;
Best Local Similarity 67.7%; Pred. No. 2.4e-164;
Matches 419; Conservative 80; Mismatches 87; Indels 33; Gaps 9;

OY 1 MASFATAAVSGRWL-----GNHTQPLSSQSSDLSYCSLP-----MASRYT 45
DB 51 MASTTLTLPSTQFLDRPTSTSSSSSRPKIQS---LSFSSLTNRKKKLIVPCYSSSVN 105
OY 46 RLAVNSAIAHT---PPALHFPKSSNSPAIYVKKPAKKSNTKOMLPPORAAAALDAE 101
DB 106 KSSVSSSSSQSPTFKPPSR--KKLCNDVTNLI-PRTTQONPK-LMPVORTAAMVDAVE 160
OY 102 GFLVSHH-KLAPLPTKADPSVOIAGNFAPVNEQPYARNLPVVGKLPDSIKGYVRNGANP 160
DB 161 NAMISHERRRHHPKTDADAVOIAAGNFPPVREKPYVHNLPVGTGYECCIGGYVVRNGANP 220
OY 161 LHEPYTGHHPFGDGMVAHVAFKFEHGSASACRFTQTNRFVQEROLGRVPYPAIGELGH 220
DB 221 LHKPVSGHLLFPGDGMVAHVAFKFEHGSASACRFTETNRLVQEROLGRVPYPAIGELGH 280
OY 221 TGIARLMLFYAARAAGIYDPAHGTGVANAGLYVFNGLRLAMSEDDLPYOVOTTPNGDLKT 280
DB 281 LGIARLMLFYAARAAGIYDPAHGTGVANAGLYVFNGLRLAMSEDDLPYHVAVKTQGDLETT 340
OY 281 VGRFPDQGLESTMLAHKRVADPESGELFALSYDVVSKPLKFFRSPOGTSKPDVEIOLD 340
DB 341 SGRYPDQGLKSTMLAHKRVADPESGELFALSYDVVSKPLKFFRSPOGTSKPDVEIOLD 400
OY 341 OPTMMHDAFAITENFVVPDQOVVFKLPEMIRGSPVYVDKNAVARFGILDKYAEDSSNIK 400
DB 401 OPTMMHDAFAITENFVVPDQOVVFKLPEMIRGSPVYVDKNAVARFGILDKYAEDSSNIK 460
OY 401 WIDAPDCFCFHLMANMEEBETDEVVYVIGSCMTPPSIFNESEDNLKSVLSEIRLMLKTGE 460
DB 461 WIDAPDCFCFHLMANMEEBETDEVVYVIGSCMTPPSIFNESEDNLKSVLSEIRLMLKTGE 520
OY 461 STRRPITISNEDQOVULEAGVNRNMLGRKTKPAYLALAEPMKVSGFAPKVDLTGGEVKKH 520
DB 521 STRRPITISNEDQOVULEAGVNRNMLGRKTKPAYLALAEPMKVSGFAPKVDLTGGEVKKH 577
OY 521 LYGNRRYGGEPFLFPGEGEEDGYILCFVHDEKTKWSELOIVNAVSLVEATVKLPSRV 580

DB 578 IYGEKYGGEPFLFPGSDDEEDGYIMVHDEEKVKSSELQILNAVNMKLXATVTLPSRV 637
OY 581 PYGHGTFTFGADLAKQVY 599
DB 638 PYGHGTFTFGADLAKQVY 656
RESULT 6
OYFS24 PRELIMINARY; PRT; 612 AA.
AC OYFS24.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NEOXANTHIN CLEAVAGE ENZYME.
GN CPRD65.
OS Vigna unguiculata (Cowpea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseolaeae; Vigna.
OX NCBI_TaxID=3917;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20317197; PubMed=10859185;
RA Iuchi S., Kobayashi M., Yamaguchi-Shinozaki K., Shinozaki K.;
RT "A stress-inducible gene for 9-cis-epoxycarotenoid dioxygenase
RT involved in abscisic acid biosynthesis under water stress in drought-
RT tolerant cowpea";
RL Plant Physiol. 123:553-562(2000).
DR EMBL: AB030293; BAB11932.1; -
SQ SEQUENCE 612 AA; 67715 MW; E3263A13293B188B CRC64;

Query Match 68.8%; Score 2168.5; DB 10; Length 612;
Best Local Similarity 68.6%; Pred. No. 3.2e-164;
Matches 423; Conservative 61; Mismatches 98; Indels 35; Gaps 8;

OY 8 AAVSGKWLGNHTQPLSSQSSDLSYCSLPMAASRVTRKLN-VSALHTPPALHFPKO- 65
DB 3 SSASNTWFMATLPSPFKDLPSTS--SPINLPL--RKTISSSTITICSLOT--LHFPKQY 56
OY 66 ---SSNSPAIYVKKPAKKSNT-----KOMNLFQRAAAAALDAA 100
DB 57 OPTSTSTATTTPTPIKTIITTTTPRETNPISDTNQRLPQKWNFLQKAAATVLDLV 116
OY 101 EGFVLSHEKHLPLPTADPSVOIAGNFAPVNEQPYARNLPVVGKLPDSIKGYVRNGANP 160
DB 117 EGFVLSHEKHLPLPTADPSVOIAGNFAPVNEQPYARNLPVVGKLPDSIKGYVRNGANP 176
OY 161 LHEPYTGHHPFGDGMVAHVAFKFEHGSASACRFTQTNRFVQEROLGRVPYPAIGELGH 220
DB 177 LHEPYTGHHPFGDGMVAHVAFKFEHGSASACRFTQTNRFVQEROLGRVPYPAIGELGH 236
OY 221 TGIARLMLFYAARAAGIYDPAHGTGVANAGLYVFNGLRLAMSEDDLPYOVOTTPNGDLKT 280
DB 221 TGIARLMLFYAARAAGIYDPAHGTGVANAGLYVFNGLRLAMSEDDLPYHVAVKTQGDLETT 340
OY 221 VGRFPDQGLESTMLAHKRVADPESGELFALSYDVVSKPLKFFRSPOGTSKPDVEIOLD 340
DB 281 VGRFPDQGLESTMLAHKRVADPESGELFALSYDVVSKPLKFFRSPOGTSKPDVEIOLD 400
OY 281 VGRFPDQGLESTMLAHKRVADPESGELFALSYDVVSKPLKFFRSPOGTSKPDVEIOLD 460
DB 341 OPTMMHDAFAITENFVVPDQOVVFKLPEMIRGSPVYVDKNAVARFGILDKYAEDSSNIK 400
OY 341 OPTMMHDAFAITENFVVPDQOVVFKLPEMIRGSPVYVDKNAVARFGILDKYAEDSSNIK 416
DB 341 OPTMMHDAFAITENFVVPDQOVVFKLPEMIRGSPVYVDKNAVARFGILDKYAEDSSNIK 460
OY 401 WIDAPDCFCFHLMANMEEBETDEVVYVIGSCMTPPSIFNESEDNLKSVLSEIRLMLKTGE 460
DB 417 WIDAPDCFCFHLMANMEEBETDEVVYVIGSCMTPPSIFNESEDNLKSVLSEIRLMLKTGE 476
OY 461 STRRPITISNEDQOVULEAGVNRNMLGRKTKPAYLALAEPMKVSGFAPKVDLTGGEVKKH 520
DB 477 STRRPITISNEDQOVULEAGVNRNMLGRKTKPAYLALAEPMKVSGFAPKVDLTGGEVKKH 535


```

0y 521 LYGDNRGKGPELFLPFGSGGDEDEYIIICFVHDEKTMKSELOIYNAVSLTEAVYKIPSRV 560
    ||:::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 536 MYGEKRGGELEFLP-NGQKEDDDIIILAFVHDEKEMKSELOIYNAONMLKLEASIKLPSRV 594
    ||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
0y 561 PYGFHGTFFICAGDILAKQ 597
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 595 PYGFHGTFFIRHSLDKQ 611
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```

RESULT	7	
Q9M6E8		
ID	Q9M6E8	PRELIMINARY;
Q9M6E8		PRT; 615 AA

DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE 9-cis-EPOXYCAROTENOID DIOXYGENASE.
GN NCED1.
OS Phaseolus vulgaris (Kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotsids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, TOP CROP;
RC MEDLINE=20079657; Pubmed=10611388;
RA Qln X., Zeevaert J.A.;
RT "The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory
RT step of abscisic acid biosynthesis in water-stressed bean.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:15554-15561(1999).
DR EMBL; AF190462; AAF26356.1.f. -
KW Dioxxygenase.
KW SEQUENCE 615 AA; 68075 MW; 0CC10F8627DE130 CRC64;

Query Match: 68.7%; Score 2165; DB 10; Length 615;

Matches 426; Conservative 59; Mismatches 95; Indels 36; Gaps 9;

```

QY      11  SGRMLGNGHNPPLSSQSSDLSYCS--  -LPMASRVTKLN-VSSALHTPEALHPK 64
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      6  SNTMI--NTLIPSSCSSEPFKDLASTSSPTTLLEPKRRSSNTNTVTCISQT--LHPK 60
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY      65  Q-----SSNPALVYKPKAKE-----SNTKQ-----MNFQRAAAALDAE 101

```

Db 61 QYQPTSTSTTTPPIKPTTTTTTPHRETKPLSDTKQPFPPQKWNFLQKAAATGLDME 120

QY 102 GFLVSEKHLPLPKTADPSVQIAGNEAPVNEQPVRRNLPVVGKLPDSIKGVYVRNGANPL 161

Db 121 TALVSHESHKHLPTADPKVQIAGNEAFVPVEHAADQALPVVGKIPKCIDGVYVRNGANPL 180

QY 162 HEPVTGHHFFDGDGMVHAVKFEHGSASYACRFQTQTNREVQEROLGRPVFPKAI GELHGT 221

Db 181 YEPVAGHHFFDGDGMVHAVKFTNGAASYACRFTETQRLAQEKSLGRPVFPKAI GELHGS 240

QY 222 GIARLMFYARAAGIVDPAHGTCVANAGLVYFNGRLLAMSEDDLPYQVQITPNGDKTV 281

Db 241 GIARLLFYARSLFQLVDGSHGMGVANAGLVYFNNHLLAMSEDDLPHYVRITSNGDLTV 300

QY 282 GRFDFDQLESTMIAPKVDPESEGLFALSYDVVSKPYLKIFRFSPDGTSKDVEIQLODQ 341

Db 301 GRYDFNGQLNSTMIAHPKLDPVNGDLHALSYDVQKPYLKFRFSADGVKSPVEIPLKE 360

QY 342 PTMHDEAITENFVVPDQVFEKLEPMIRGGSPVYDKNKVAREGILDKYAEDSSNIKW 401

Db 361 PTMHDFAITENFVVPDQVFKLTETLTGGSPVYDKNKTSREGILDKNAKANAMRW 420

QY 402 IDAPDCFCFHLNNAWEEPETDEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNKTGES 461

Db 421 IDAPECFCFHLNNAMEPETDEIVIGSCMTPADSI FNECDESLKSVLSEIRLNLRTGKS 480

```

QY 462 TRRPITSMEDQVNLNACAMVRNNLGRKTRFAYTALAEPMPKVSGFAXVLLTGEVKKH 521
Db 481 TRRPITTS-DAAQVNLNACAMVRNNLGRTOFAYTALAEPMKVSGFAXVLLFSGEYWK 539
QY 522 YGDNNYGEPEPLFEGEGEEDGYILCPVHDEKTMKSELOVNVANSLEVEATYKLPSPRV 581
Db 540 YGEKFGGEPEPLF-NGEEBGDGYILAVHDEKEMKSELOVNAVNLKLEASTIKLPSPRV 598
QY 582 YGFHGTFTGADDLAKO 597
Db 599 YGFHGTFTHSKDLRKO 614

```

RESULT	8	
Q9AXZ4		
ID	Q9AXZ4	PRELIMINARY;
		PRT;
		625 AA

DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)
DE 9-CTS-BPOXYCAROTENOLD DIOXYGENASE.
GN NCED3.
OS *Persea americana* (Avocado).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
OX NCBI_taxid=3435;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LULA.
RA Chernys J., Zeevaert J.A.D.;
RT "Abscisic acid in avocado fruit";
RU Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF224671; AAK00623.1; --
DR InterPro, IPR000221; Protamine_P1.
DR PROSITE, PS00048; PROTAMINE_P1; UNKNOMN.1.
KW Dioxigenase.
SQ SEQUENCE 625 AA; 69724 MW; A56F3042D5F2CB6 CRC64;

Query Match	65.68; Score 2066.5; DB 10; Length 625;
-------------	---

Best Local Similarity	Seq. NO.	Seq. ID	Seq. Length	Seq. Type	Seq. Source
Matches	392	Conservative	74	Mismatches	93
Indels	45	Gaps	6		

QY 38 LPMASRVT RKLNVSSALHTPP-----ALHFP-----KQSS 67

Db 22 LPISKNLSPKNEIMLKHNTPLIQCCSHSPSSSAAVLHLPKQPTKSKPSIKKGEKST 81

68 NSPAIVVKPKAKESNTKQ-----MNLQRAAAALDAEGFLVSH--EKLHPLPKT 116

Db 82 LTPSIEKNPGSHQVKTQSGPNRVGPNWNIQRTAAAFALDAIEKLIARVLERRHPPLPKT 141

QY 117 ADPSVQIAGNEAPVNEQPVRRNLPVVGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGM 176

Db 142 ADPEVQIAGNEAPVAEHVPQHGIPVAGRIPRCLDGVYVRNGANPLEPIAGHHFFDDGM 201

QY 177 VHAVKEHGSASYACRFTQTNREVOEROLGRPVFPAIGELHGTGIARLMLFYARAAG 236

Db 202 IHAVERNGSASYSCTETRLVQERQLSRPIFPKAIGELHGHSGIARLLLFYTRGLEG 261

QY 237 IVDPAHGTCVANAGLVYFNGRLLAMSEDDLPYQVQITPNGDKTVGRFDEGQLESTMIA 296

Db 262 LVNADEGMVANAGLVYFNRRLLAMSEDDLPYHVRITPPSGDLKTVGRHDEFDNLRRSSMIA 321

QY 297 HPKVDPESEGEFALSVDVSKPYLKYFRFSPDGTSPDVEIQLDQPTMMHDAITENFV 356

Db 322 HPKLDPESGELFSLSYDVARKPYLKYFHFAPDGWKS PDVEIPLDRPTMIHDEAITENFV 381

QY 357 VPDQVVEKLEPMIRGGSPVYDKNKVARFGILDKYAEDSSNIKWIDAPDCFCFLWNW 416

Db 382 IPDQVVFKEEMIRGSPVYDKNTKTSRFGILPKYADASEMIWVADPDCFCFHLNNAW 441

QY 417 EEPETDEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNKTGESTRPIISNEDQVNL 476

Db	442	EEPESEGEVVVVVSGCMTTPDPSJFNENNESIASILFEINLNRTGSTRRTII-DPQKPLNL	5000
Qy	477	EAGVYNNRMKCRKTKTFAYLALAEWPWPVYSGFAKVDLTTEGYKKHLXGDNRYGSEBPLFLPG	5366
Db	501	EAGWYNNRRRLTKRTRTFAYLALAEWPWPVYSGFAKVDLTGTEGYNRRFVYGERQFGGEPYPIR	5600
Qy	537	E---GGDEDEGIIICFVHDEKTKMSELIQIVAAVSLAEVATYKPLSRPYVFGHFGTIGADD	5933
Db	561	EPSTSGREDDEGVVVSFMHDEKTSKSEILLIANMNRLEASVMLPSRVPYVFGHFGTISR	6200
Qy	594	LAKQ	597
Db	621	LAKQ	624

RESULT	9
ID	09C621
AC	09C621
DT	01-JUN-2001 (TREMBLrel, 17, Created)
DT	01-JUN-2001 (TREMBLrel, 17, Last sequence update)
DT	01-JUN-2001 (TREMBLrel, 17, Last annotation update)
DE	9-CIS-EPOXYCAROTENOID DIOXYGENASE, PUTATIVE.
GN	2H7.10.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Euxaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX	NCBI_TaxID=3702;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV, COLUMBIA.
RX	MEDLINE=21016719; PubMed=11130712;
RA	Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA	White O., Alonso J., Alfati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA	Buehler E., Chan A., Chao Q., Chen H., Cheuk R., Chin C.W.,
RA	Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA	Dunn P., Etgu P., Feldidym T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA	Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
RA	Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA	Kim C.J., Koo H.L., Kremetatskaja I., Kurtz D.B., Kvan A., Lam B.,
RA	Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA	Liu X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA	Milletscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA	Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA	Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA	Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA	Ueberback T., Van Aken S., Vaysberg M., Vysotskaja V.S., Walker M.,
RA	Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT	"Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT	thaliana.";
RL	Nature 408:816-820(2000).
DR	EMBL, AC074176; AAC50855.1; -.
KW	Dioxygenase.
SO	SEQUENCE 589 AA; 65336 MW; 6A79B2BF7EAA8179 CRC64;
Query Match	64.9%; Score 2044.5; DB 10; Length 589;
Best Local Similarity	68.1%; Pred. No.2.3e-154;
Matches 390; Conservative 69; Mismatches 101; Indels 13; Gaps	
0Y	27 SOSSDLSYCSSLPMASSRYTRKLTANVSALHPHPALHPFKQSSNPATVYVKKAKESNTKOM 86
DB	27 SPSSSVSTNTNPKR-----RRKLSSANVSVDIPNLNLFNPYSPNPII-----PEKDTSRW 76
0Y	87 NLFORAAAAALDAAGFLVSHSEKLNHPKTAADPSVQIAGFNAPVNEQPVARNLPVGVKKP 146
DB	77 NPLORAAASALDFAETALLRERSKRPDKPTVDPRHQISGNVAPBEQSVKSSLSVDGKIP 136
0Y	147 DSIKGVYVRNANGNPLHEVYVTHGHFFDDGGGMAVNAVKFEHGSASVACRTQTQNNRQVQKQLG 206
DB	137 DCIDGVYVRNANGNPLFEVYVSHHLEFDGGGMAVNAVKINGASVACRTETTERLVQKQKQLG 196

Qy	207	RPPKPAIGELHGHTGIALRLMEFARAAGAVDPAHCTGYANGLYVENORILLAMSSDDL	2669
Db	197	SPIFPKAIGELHGHSGLARLMLFARGLFGLLNKNGTGAANAGLYFHHORLLAMSSDDL	2566
Qy	267	PYVOVITPENGDLKTVYGRFDEPGQLESTIMIAHPKVDPESGELEFALSTOVNSKPYLKYERFS	3368
Db	257	PYVARYVDNQGDELITIGRDFPDGQLSAMIAHPKIDPYTKELFALSTOVNKKPYLKYERFS	3166
Qy	327	PDGTRKSPDEVTQLODPYTMHMDFAITENFVVVPPDOOVVFKLEPMIRGSGPVVYDKNKVAPF	3866
Db	317	PEGEKSPPEVETPLASPTMHMFATITENFVVVPPDOOVVFKLSDMELFGSPYKYOGEKISRF	3766
Qy	387	GILDKYAADSSINIKWIDAPDFCCHILNANWEEPEPTDEVYVYTGSCMTPPDISFNESDNLK	4466
Db	377	GILPNNADASSEMVAWSEPEFFCEHILNANWESPETDEVVYVYTGSCMTPADDISFNESCDQLN	4366
Qy	447	SVLSSEIRLNLGTGESTRPIISNDDOQVNLGAVVNNMLGRKTKFAYLALAEWPWPVSG	5066
Db	437	SVLSSEIRLNLKGTGSTRPIIPG-SVQNNLEAGVYNNRLLOKTRKYAYLALAEWPWPVSG	4955
Qy	507	FAKYVDLTTGVEYKHLXGDNRYGGEPLFLPG--EGGEDEGYILCFVHDEKTMWSELOIVN	5644
Db	496	FAKYVDLSTGEVKNHFYGGKRYGGEPPFLPRGLJESDEGDDGYIMSFVHDESEWSELIHVN	5555
Qy	565	AVSLEVEATVKLPSRVPYGFHGTITGADDLAKQ	597
Db	556	AVTLELEATVKLPSRVPYGFHGTITGVNSADMLNQ	588

```

RESULT 10
049505 PRELIMINARY; PRT; 583 AA.
AC 049505;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NEOXANTHIN CLEAVAGE ENZYME-LIKE PROTEIN.
GN F28J12.10 OR AT4G18350.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bewan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Mayer K., Scheller C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
EU Arabidopsis sequencing project;
RL submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL021710; CAA16715.1; -.
DR EMBL: AL161548; CAB78837.1; -.
DR InterPro: IPR004294; RPE65.
DR Pfam: PF03055; RPE65; 1.
SE SEQUENCE 583 AA; 65066 MW; A138P93542E50852 CRC64;

```

Query Match	63.2%	Score 1991;	DB 10;	Length 583;
Best Local Similarity	64.5%;	Pred. NO. 4.1e-150;		
Matches 380;	Conservative 82;	Mismatches 105;	Indels 22;	Gaps 6

Db	11	GGIKTW----	QAQIDGFGRIRKQPRVICTQYIDVTE-LTKKROLFTPTRTATPP----	62
Oy	74	VKPKAESNTKQNNLFDORAAAAALDAAEGLVSLSEHKTLHPKPTADPSVOIAGNPADNEO	133	
Db	63	-----OHNPRLRLIFPKAAALADAAERLALISHEODSPKPTADPVOIAGNVSPVES	116	
Oy	134	PVRRLPVVGLPDSISGVVVRNCAANLPHVEYGHFEDDDGWHVAFPHGSASVACRP	193	
Db	117	SVRRLLIVYEGTIPPCIDBVTIRNCAAMPEETAGHILFDDGWHAAKLTNGSNASTACRP	176	
Oy	194	TQTNKFEVEROLGRBPVEPKAIGELHGHGTGIARLMLFYARAAGIVDPAGHGTGVANAGLVY	253	
Db	177	TKTERLVQEKRLGRBPVEPKAIGELHGHGTGIARLMLFYARGCGLINNQNGVVAANGLVY	236	
Oy	254	ENGRLAMSEDDLYQVQITPNGCLKTVYGRDPFGQLESTMIAPHKVPESGELFALSYD	313	
Db	237	FNNRLLAISEDLYOLKIITQGLQTVGRYDFDQGLKSMIAHPKLDIPYKELHALSYD	296	
Oy	314	VYSKYLKFFRFPSPDGKSPVEIQLQOOPFMHDFALTENFVVYDDQVVFYKPREMIRGG	373	
Db	297	VYKRPYLYKFFRFPSPDGKSPLELEPPLPHIHDFALTENFVYIPDQVVFYKIGEMISGK	356	
Oy	374	SPVYVDKKNVAFGLDKYAADSSNIKWIDAPDCFCEHLNAMEEPTDEVYVIGSCMP	433	
Db	357	SPVVEDEGKVSRLIMPADATEASQIITWVNSPEFCFHLNAMESPETEELIYVIGSCMP	416	
Oy	434	PDSIFNEDEMLKCVLSIRLINTKGTSESTRPIISNDDOYNLEAGVNNMGLGRKTKFA	493	
Db	417	ADSFINEDESLKSVLSIRLINTKTRTKTRSLLVND--VNLEIGVNNRNLGRKTRFA	474	
Oy	494	YLALAEPPKVSAGKAVDLYTGEVKKHLGDNRYGGEPLFLPEGEG---EDEDGYTLCP	549	
Db	475	FLAIAVPPKVSAGKAVDLYTGEVKKHYIYGEKKGGEPFLPGNSGGEENEDGYTLFCH	534	
Oy	550	VHDEKTKMSELQIVNANVSLVEAFVKLPSRYVYFGHGTFGADDLAKOV	598	
Db	535	VHDEKTKMSELQIINAVMLKLEATIKLPSRYVYFGHGTFFVDSNELYDOL	583	
RESULT 11				
Q9AXJ3	ID	09AXJ3	PRELIMINARY;	PRT; 569 AA.
AC	AC	09AXJ3;		
DT	DT	01-JUN-2001 (TReMBLrel, 17, Created)		
DT	DT	01-JUN-2001 (TReMBLrel, 17, Last sequence update)		
DE	DE	01-JUN-2001 (TReMBLrel, 17, Last annotation update)		
GN	GN	9-CIS-EPOXYCAROTENOID DIOXYGENASE.		
OC	OC	NCEID1.		
OC	OC	Persea americana (Avocado).		
OC	OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	OC	Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.		
OX	OX	NCBI_TaxID=3435;		
RN	RN	[1]		
RP	RP	SEQUENCE FROM N.A.		
RC	RC	STRAIN=CV, L01A;		
RA	RA	Chernys J., Zeevaert J.A.D.;		
RT	RT	"Abcisic acid biosynthesis in avocado fruit.";		
RL	RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.		
KM	KM	EMBL: AF224672; AAK00632.1; -.		
SO	SO	Dioxygenase.		
		SEQUENCE 569 AA: 63078 MW: CD12075EE9022461 CRC64;		

[illegible]

```

Db      65 VALDVEDESFISGVLEKRNHPLEPKTSDPAVOYSGNFARVYDENHPQNHLPVSGRIPECLDGV 124
OY      153 YVRNGANGLHEPVYGNHFEFDGDGVNVAKTEHGSASACRFOTNREVOEROLGRVPEPK 212
Db      125 YLRNGANGLLEPVAGNHFFEGDDGVNHSVLSRGTAASYACRFETETNRLVORRAIGRPVEPK 184
OY      213 AIGELHGTGTARLMLFYVARAAALIVDPRAHSTGVANNGVLYFNRRLAMSEDDLPVOYOY 272
Db      185 AIGELHGHSGIARLLLEFLFARATGALVGGSSGTGATGVLVFNHLLAMSEDDLPVHVR 244
OY      273 TPNGDLKTVGREFDEFGOLESTMIAPKVPDEPSEGLFALSTDVVSKPLKFRFSPGATKS 332
Db      245 TSSGDLLEVGRFDEFGOLNSAMIMHPKIDPASPGLFALSTVYIKKPLKFRKFTSGOKKS 304
OY      333 PVEVLIDLOPTMHDEFAITENFVVVPOQVVFKLPEMI RGGSPVYVDKNVAREGLIDKY 392
Db      305 PVEVPIPOPTMIHDEFVITENFVLIIPQOVVFKLQEMI RGGSPVYVDKKRTARFGLILKT 364
OY      393 AEDSSNIMWIDAPOCFCNHLNAMEEPEDTGVVYIGSCMTPPOSIFNESDOKKSVLSEI 452
Db      365 AADSSNGLKMWIDAPDCFCNHLTAMEEPEDTGVVYIGSCMTPPOSIFNESDOKKSVLSEI 424
OY      453 RLNTKTEGSTRAPLISNDDQOVNLEAGVWNNMIGRKTCAFYALALAPMKVSGFAKVDL 512
Db      425 RLNTKLTGSSRREI--DPSRLNLEVGWVNNNRNLRGRBPVYSLSHCHCRPMKVKVSPAKVDL 482
OY      513 TTGEVKKHLVGDNRKYGGAPFLPCGBG-GEEDEGYILCFVHDEKTKWSELOIVAAVSLAEV 571
Db      483 STGEVTKFYEGOCYGGEPYEVSRDPVAPAEEDDGVVLSFMHDEKTAARSELLIVAINATWOLE 542
OY      572 ATVKLPSRVYGFHGTFIGADILKO 597
Db      543 ASVKLPSRVYGFHGTFIGADILKO 568

```

Query Match	61.3%	Score 1930	DB 10	Length 604
Best Local Similarity	63.3%	Pred. No. 3.2e-145		
Matches 375	Conservative 73	Mismatches 122	Indels 22	Gaps 8
QY	22	PLSSQSSDLSYSSSLPMASRYTRKLNYSALHP	---PALHFPKSSNSPAIV	----74
		:::::	:	:::::
Db	17	PASRRARAS----	NSVRFSPRAVSSVPPAECLQAFHFKPVADLPAPSKRPALAIAPGHA	71
QY	75	-PKAKESNTKQNLFFQRAAAALDA-AGCELYS	-HEKIHLPKTDPSVQIAGNAPVN	131
		:::::	:::::	:::::

Db	72	AAPRAEGCKQKQLNLFQRAAALADLAFEEGFAVNI	ERHGLPSPADAPVQJAGNFAVVG	133
Qy	132	EOPRRNLIPVYVGLKIPDSIKGVYVRANGANPL	AHEVYTHGHEFFDGDGMVHAFEGHSA-SYA	190
Db	132	ERPPEHLEPVSGRIIPRIDIVYVRANGANPC	DFDVAHGHLEFDDGMVHALRLINGNAESYA	191
Qy	191	CRFQTNFVEVEROLDGPRVPKAIIGELHGHGTG	FAIRLMLPYARAAGIYDPAPGTCVANAQ	250
Db	192	CRFETIARLROERAIIGRPVPKAIIGELHGHST	IAIRLALYARAAGIYDPASAGTCVANAQ	251
Qy	251	LYVFNGLLAMSDEDDLPYQVQITPNQDLKTVR	GFDFDGLJESTMTIAHPKVPESGELFAL	310
Db	252	LYVFNGLLAMSDEDDLPYHVRVADDDGLETVGR	YDERDGLGCMALAHPLRDPATGELHAL	311
Qy	311	SYDVASKPYLYKTFEFSDDGKTSDEVEILOD	PMPMDPRAITENFVYVDPQOYVEFLPMT	370
Db	312	SYDIKIRPYLYKTFEFSDDGKTSDEVEILOD	PMPMDPRAITENFVYVDPQOYVEFLPMT	371
Qy	371	RGSFVYVYDKKVAARFEGILDKYVADSSNTK	KWIDAPCCPFHLMNAWBEPTDEVVYISGC	430
Db	372	RGSFVYVYDKKVAARFEGILDKYVADSSNTK	KWIDAPCCPFHLMNAWBEPTDEVVYISGC	431
Qy	431	MTPEPSIFNESDENLKLVSLEIRLNLKLGEST	RRPIITSNEDQVNLVLEAGVVRNMLGKRT	490
Db	432	MTPEPSIFNESDENLKLVSLEIRLNLKLGEST	RRPIITSNEDQVNLVLEAGVVRNMLGKRT	490
Qy	491	KFATIALAEPPPKVSGAKAVDLTTGVEKKHL	IGDNRNGEPLFLPEEGGE----	EDBGY 545
Db	491	KFATIALAEPPPKVSGAKAVDLTTGVEKKHL	IGDNRNGEPLFLPEEGGE----	EDBGY 545
Qy	546	ILCEVHDEKTKWKSLEQIVNVSLEVEATVYLP	FRVRYGFGHGFITGADPLAKO	597
Db	551	VLTIVHDERAGTSELVLAHADRILEATVQLPS	RVPFGHGFITIGQLELKO	602
RESULT 13				
QyLRM7	09LRM7	PRELIMINARY; PRT: 577 AA.		
AC	09LRM7	PRELIMINARY; PRT: 577 AA.		
DT	01-OCT-2000 (TReMBLrel. 15, Created)			
DT	01-OCT-2000 (TReMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TReMBLrel. 15, Last annotation update)			
DE	9-CIS-EPOXYCAROTENOID DIOXYGENASE.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.			
NCBI	TaxID=3702;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-COLUMBIA.			
RA	Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;			
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
RP	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-COLUMBIA.			
RX	MEDLINE=20277480; PubMed=10819329;			
RA	Nakamura Y.;			
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence			
RT	features of the regions of 4,504,864 bp covered by sixty P1 and TAC			
RT	clones.";			
RL	DNA Res. 7:131-135(2000).			
DR	EMBL; AB028621; BAB01363.1; .			
KM	Dioxygenase.			
Qy	SEQUENCE 577 AA; 63821 MW; 515E0A1B321B4FE1 CRC64;			

Query Match	52.8%;	Score 1663.5;	DB 10;	Length 577;
Best Local Similarity	55.4%;	Pred. No. 5.2e-124;		
Matches 323;	Conservative 137;	Mismatches 100;	Indels 23;	Gaps 9;

QY 24 LSSQSSDIYCCSLPMASRYTRKLNWSS-ALHTPPALHPKQSSNSPAIVVKPKAKESN 82

```

Db 9 LBPKTSRSHLLDQPKKANISRIILINPKIPIPLDPLTSPVSPS-----VALKPTPYM 62
QY 83 TKOMNLFORAAAALDAAE-GEIVSHEKHLPLEKTDAPSVQIAGNEAPVNEOPVRNRLPY 141
Db 63 ---INTLQKLAATMLDKIESSIIVIPMEQONPLPPTDPAVLGSNENFPVNECEVONGLEV 119
QY 142 VCKLIDPDSIKGVYVNGANPNLHERPTGHNHFDGOCMVAAV--BEHGSASACFPFTQNR 199
Db 120 VQIIPSCICKGYIINGANPMPRLAGHHLFDGDMHIAVIGPDN-OVYSCHRYTKTNRL 178
QY 200 VQERLGRPVPAKAIGETLHGHTGJARLMLFARAAAGIVPAHAGTGAANAGLVFNGRLL 259
Db 179 VQERALGHSVFPKPIGELHGHSGIARLALFTPARAGISLVDTGKMGYANAGVFPNGRLL 238
QY 260 AMSEDDLEPYOVITPNGLDKVTGRFDEGOLESTMIAPHKVDSEGEFLASVYNSKPY 319
Db 239 AMSEDDLEPYOVKIDGGGLDETIGRFEGDDQIDSSVIAHPKVDATGTDLHTLSYVNLKPH 298
QY 320 LKYREFSPDGTKSPDVELQDLPIMHDEFAITENFVVVPOQVYFKLEPIRGSPRYVD 379
Db 299 LRYLKFNFCKGTADVETLPEPTIMHDAITENFVVIPOQVYFKLEPIRGSPRYV 358
QY 380 KNKYARGLILDKVYADESSNIKWIDAPCCFCHILMANAEBEPDE---VVVITSCMPRP 435
Db 359 KEKARRGVLSKQDPLTGSIDIMWVDPCCFCHILMANAEB--RTGEGDPIYVITDSCMSPD 417
QY 436 SIFNESEDENLKSVEISEIRLNLKTGESTRPPIIENEDQVNLDEAGVYVNNMIGKRTFAYL 495
Db 418 TTFSSGEPTRVELSEIRLNNRTKESRKVIYVG---VALLAGHINRSYVGRKSQFVYI 473
QY 496 ALAEWPVKVSEFAVVDLTGTGEVKKHLTGDNRYGCEPLFLRGEGEEDEGYIILCFVHDEKT 555
Db 474 AIADPWPKCSGIAVVDIONGVSEFNENGRSPRFGCEPCEVEGEGEEDKGYMGVGFVRDEEK 533
QY 556 WKSELOIYNAVSLAEVAFVTKLPSVPRVPGFHTGFGADDLAKOV 598
Db 534 DESERVVADDMKQVANAIRLPEKVPVIGFHTVSEMOLEOV 576

```

RESULT	14		
ID	049675		
AC	049675	PRELIMINARY;	PRT; 595 AA.
DT	01-JUN-1998	(TREMBLrel. 06, Created)	
DT	01-JUN-1998	(TREMBLrel. 06, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	NEOXANTHIN CLEAVAGE ENZYME-LIKE PROTEIN (AT4G19170/118B16_140).		
GN	118B16.140 OR AT4G19170.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_Taxid=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Bevan M., Benes V., Rechmann S., Borkova D., Ansoerge W., Bancroft I.,		
RA	Mewes H.W., Mayer K., Schueller C.;		
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Benes V., Rechmann S., Borkova D., Ansoerge W., Mewes H.W., Lemcke K.,		
RA	Mayer K.F.X.;		
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
RP	[3]		
RP	SEQUENCE FROM N.A.		

RA Cheuk R, Chen H, Kim C.J., Koesema E, Meyers M.C., Bahh J.,
 RA Bowser L, Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Newman G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,

RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL021687; CAB16706.1; -
 DR EMBL; AL161550; CAB78919.1; -
 DR EMBL; AY056789; AAL10480.1; -
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65; 1.
 SQ SEQUENCE 595 AA; 65601 MW; 711EDC4AF59920B9 CRC64;

Query Match 30.7%; Score 968; DB 10; Length 595;
 Best Local Similarity 36.7%; Pred. No. 1.5e-68;
 Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;

QY 42 SRVTRKLVSSALHTPPALHPKSSNSPAIVVKKAKESNTKQNLFORAAALDAE 101
 DB 42 SPITNPSQNNDRNRKPKTLH---NRTNHTLVSSPKLRPEMTLATALF---TIVEDYIN 94
 QY 102 GFLVSHKHLPLKPTADSVQIAGNFARVNDQPVARNLPV-GKLPDSIKGVYVNGANP 160
 DB 95 TFDIDP-----FSRPSVDKRVHLSDFNFAPVLDLPTDCEIIGHITPLSLNGAYIRNGNP 149
 QY 161 LHEPVYTHGFDDGDMVAHVAFKFGSASACRFOTQTNRFQEROLGRVFPVPAKIGELGH 220
 DB 150 QFLPRGPHYLEDGDMHAIKIHKNKATLCSRYKTYKYNENKQKAGAYMPVFSGFNGV 209
 QY 221 T-GIARLMFYARAAAGIVDPAHGTGVANAGLVYFNGRILAMSEDDLEPYQVITPNDLK 279
 DB 210 TASVARGALTAARVLTGQYNPNVNGIGLANTSLAFPSNRLFALGESDLPYAVALTESDIE 269
 QY 280 TVGRDFEGOLESTMIAPKVPDESGELFALSYDVNSKPYLKYFFSPDGTKSPVEI-Q 338
 DB 270 TIGRDEFGKGLAMSTAHKPTDPTIGETFAFRYGPV-PPFLTYFFDSAGKKQROVPFIS 328
 QY 339 LDQPTMHDFATITENFVVVPPDOQVVF--LPEMIRGSGPVYVDNKKVAFSGIILDKYAE 395
 DB 329 MTSPEFLHDFATITRNAHFAELQGLGRMMMLDYLEGSSPYOTDNGKTPRLGLVITKYACD 388
 QY 396 SSNIKWIDAPDPCFHLNNAWEEPTDEVVIGSCMPPTDIFNSDENLKSIVSEIRLN 455
 DB 389 ESEMWFEPVPGFNILHAINADEDDGNSVLLAPNMSIEHTLERMD-LVHALVEKVKID 447
 QY 456 LKTGSTPRPIISNEDQVNLKAGVNNRMGKRTKRFAYLALAEWPVYSGAKVDLTG 515
 DB 448 LVTGIVRRHNPISAR---NLDFAVINPAFLGRCSRYAAAGDMPKISGVVAKLDVSKG 502
 QY 516 E-----YKHLHYGDNRYGGEPLFLPREGS---EEDEGVIICFVHDEKTKWSELDIVNAV 566
 DB 503 DRDQCTVARRKATGSCCYGGEPPFVARDPQNPAAEEDDGVVITYVHDEVTEGSEKFLVMDAK 562
 QY 567 S--LEVEATVKLPSRVYPYGFHGTFIGADDLAK 596
 DB 563 SPELEIYAAVRLPRRVYPYGFHGLFYKESDLNK 594

RESULT 15
 Q94IR2 PRELIMINARY; PRT; 543 AA.
 AC Q94IR2;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CAROTENOID 9,10-9',10' CLEAVAGE DIOXYGENASE.
 GN CCD1.
 OS Phaseolus vulgaris (Kidney bean) (French bean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosoids 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
 OC NCBI_taxid=3885;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Schwartz S.H., Olin X., Zeevaert J.A.D.;
 RT "Characterization of a Novel Carotenoid Cleavage Dioxygenase from
 RT plants";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY029525; AAK38744.1; -
 KW Dioxygenase.
 SQ SEQUENCE 543 AA; 61100 MW; F22C9883A05325F7 CRC64;

Query Match 30.5%; Score 959.5; DB 10; Length 543;
 Best Local Similarity 38.2%; Pred. No. 6.3e-68;
 Matches 208; Conservative 99; Mismatches 181; Indels 57; Gaps 14;

QY 98 DAEEFLVSHKHLPL-----KTAD-----PSVQINGNAPVNDQVVR 136
 DB 8 NGAEGGLV--KVDPKPNFGSSKVIDLEKLLVLFYDSSLPHHYLTGNFGPVTEPTPT 64
 QY 137 RNLPVYGLKPDISIKGVYVRNGANPLHEPVYTHGFDDGDMVAHVAFKFGSASACRFOT 196
 DB 65 KDLPRYKHLPRCLNGEFPYRVGNPKFARVAGTHWFDGDMHIGLRKDKGKATYVSREYET 124
 QY 197 NFEVRODLGRPVPRKAIAGELHGTGIARLMFYARAAAGIVDPAHGTGVANAGLVYENG 256
 DB 125 SRLKQEEYFGSKFMYK-IGDLKGLGLMVNIMHMLRTKLKLVLDLSYGGGTTTALVYHHG 183
 QY 257 RLAMSEDDLPYQVITPNDLKTYGRDFDQGLSESTMIAPKVPDESGELFALSYDVS 316
 DB 184 KILALSEADKRPVATKVEFGDQDTGLMDYDKRLGHSFTAHKVPDFTGEMFSFGY-AHT 242
 QY 317 KPYLYFFRSPDGTSPVVEIQLDQPTMHDFATITENFVVVPPDOQVVFKLPPEMIRGSGPV 376
 DB 243 PLYITRYVYSKGVYHNDVPRTISDPIIMHDFATITENAVFMDLPLIRPKMVAKNKTLI 302
 QY 377 V-YDNKVAAREGILDKYAEDESSNIKWIDAPDPCFHLNNAWEEPTDEVVIGSCMPRP 434
 DB 303 FFSFDSTKARFGLVLRPRYAKDEQHIWFFELPNCFIHNNANWE--BEDEVVLTICRLQPK 360
 QY 435 -DSI--FNESDENLKSIVSEIRLNLKTGESRRRIISNEDQVNLKAGVNNRMGKRT 490
 DB 361 LDNVGTYQVKLENFSNELYEMKFMKTEGASOKKLAS--TVDPPEVVENTTGRKQ 415
 QY 491 KRAYLALAEPMKVSAGFAKVD-----LTGGEVYKHLV--GDNRYGGEPLFL-- 534
 DB 416 RYVGTITDSTIAKVGIIKFDLHAPRDHCKELEVGANVGILYDLGPKGFSEAVYIRV 475
 QY 535 PEGGEDEEGYILCFVHDEKTKWSELDIVNAVSEVE--ATVKLPSRVYPYGFHGTFIGAD 592
 DB 476 PGEISEEDDGYLVLVHDEBNACKSFVHYIDAKTMSADPAVAVELPNRVYPYGFHAFVTEE 535
 QY 593 DLAKQ 597
 DB 536 QLOEQ 540

Search completed: July 19, 2002, 10:15:43
 Job time: 235 sec

-

Claim 3; Fig 9; 101pp: English.

The invention relates to neoxanthin cleavage enzymes and their corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key role in endogenous abscisic acid (ABA) biosynthesis under drought stress. Neoxanthin cleavage enzyme is used for improving stress tolerance in a plant when expressed in a plant cell. The invention also relates to methods for increasing or decreasing stress tolerance in a plant by introducing the DNA into the plant, and a transgenic plant into which a neoxanthin cleavage enzyme is introduced. The improvement of stress tolerance in plants is useful, for example in plant breeding. Neoxanthin cleavage enzyme genes are useful for producing transgenic plants. An arabinid can be improved by growing transformant weed for several years and then removing the weed by specifically lowering stress tolerance in the weed by inducing an inducible promoter. The present sequence is Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCD3 protein. The AtNCD3 cDNA is obtained from an Arabidopsis plant-derived cDNA library using a cDNA of the CPR65 (Cowpea Responsive to Densitration) gene isolated from cowpea plant as a probe.

Query Match	100.0%	Score 3150;	DB 22;	Length 599;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 599; Conservative	0;	Mismatches	0;	Gaps 0;

Qy	1	MA\$FATTAASGRMLGGNHOPPLISSOSSDL\$VCS\$LPMA\$VTRKLN\$SALH\$PPL	60
Db	1	masftaacaavgrwlgqnhuppi\$ssqssdls\$ycs\$lpma\$vt\$rkln\$sa\$lh\$pp\$ll	60
Qy	61	HF\$KOSNSP\$AI\$VVK\$PK\$AKESNT\$KNML\$F\$R\$AA\$AA\$AL\$DA\$EG\$LV\$SHEK\$H\$P\$K\$T\$ADS	120
Db	61	hfpqssnspp\$iv\$vk\$pk\$ak\$entk\$gm\$ll\$fg\$raaaaa\$lda\$eg\$lv\$sh\$ek\$h\$pk\$tk\$ads	120
Qy	121	VQ\$AGN\$FAP\$VNEQ\$PVR\$RNL\$PV\$VK\$LPDS\$IK\$GVYVR\$NG\$AN\$P\$L\$HE\$P\$VT\$G\$H\$F\$ED\$G\$M\$V\$H\$AV	180
Db	121	vq\$ag\$nfap\$veq\$pv\$rn\$lp\$vk\$lpds\$ik\$gv\$yr\$ng\$an\$pl\$he\$pv\$tk\$hf\$td\$g\$mv\$hav	180
Qy	181	KFE\$G\$NS\$V\$AC\$R\$F\$OT\$R\$F\$VQ\$E\$R\$Q\$D\$G\$R\$P\$V\$K\$AI\$G\$E\$H\$G\$T\$G\$I\$AR\$LM\$L\$F\$AR\$AA\$G\$IV\$P	240
Db	181	kfe\$gs\$ns\$v\$ac\$r\$f\$ot\$r\$f\$vq\$e\$r\$q\$d\$g\$r\$p\$vk\$ai\$g\$e\$h\$g\$t\$g\$i\$ar\$lm\$l\$f\$ar\$a\$a\$g\$iv\$p	240
Qy	241	AH\$G\$V\$N\$AG\$LV\$Y\$EN\$GR\$LL\$M\$A\$E\$D\$D\$L\$Y\$Q\$V\$Q\$T\$P\$N\$G\$L\$K\$V\$G\$R\$F\$D\$P\$G\$O\$L\$E\$S\$T\$M\$IA\$H\$P\$K\$V	300
Db	241	ah\$g\$v\$n\$ag\$lv\$y\$en\$gr\$ll\$m\$a\$e\$d\$d\$l\$y\$q\$v\$q\$t\$p\$n\$g\$l\$k\$v\$g\$r\$f\$d\$p\$g\$o\$l\$e\$s\$t\$m\$i\$a\$h\$p\$k\$v	300
Qy	301	DPE\$G\$E\$F\$AL\$SY\$DV\$V\$SK\$P\$Y\$K\$L\$Y\$F\$E\$S\$P\$D\$G\$T\$K\$S\$P\$D\$V\$E\$I\$Q\$D\$P\$T\$M\$H\$D\$F\$AT\$E\$N\$F\$V\$V\$P\$D\$Q	360
Db	301	dpe\$g\$e\$fa\$ls\$y\$dv\$v\$sk\$p\$y\$k\$l\$y\$f\$e\$s\$p\$d\$g\$tk\$s\$p\$d\$v\$e\$i\$q\$d\$p\$t\$m\$h\$d\$f\$at\$e\$n\$f\$v\$v\$p\$d\$q	360
Qy	361	QV\$R\$K\$L\$P\$M\$T\$G\$S\$P\$V\$Y\$D\$K\$N\$K\$V\$A\$R\$E\$I\$D\$K\$V\$A\$E\$D\$S\$N\$T\$M\$ID\$A\$P\$C\$F\$C\$H\$L\$M\$A\$N\$E\$E\$B	420
Db	361	qv\$r\$k\$l\$p\$m\$t\$g\$s\$p\$v\$y\$d\$k\$n\$k\$v\$a\$r\$e\$i\$d\$k\$v\$a\$e\$d\$s\$n\$t\$m\$i\$d\$a\$p\$c\$f\$ch\$lm\$a\$ne\$e\$b	420
Qy	421	TDE\$V\$V\$I\$G\$S\$C\$M\$T\$P\$D\$S\$I\$F\$N\$E\$S\$E\$D\$U\$K\$V\$L\$E\$I\$R\$U\$N\$L\$K\$T\$E\$S\$T\$R\$P\$I\$S\$N\$D\$O\$O\$V\$N\$E\$A\$G\$H	480
Db	421	td\$e\$v\$v\$i\$g\$sc\$m\$t\$p\$d\$s\$i\$f\$n\$e\$s\$e\$d\$u\$k\$v\$l\$e\$i\$r\$u\$n\$l\$k\$t\$e\$s\$t\$r\$p\$i\$s\$n\$d\$o\$o\$v\$n\$e\$a\$g\$h	480
Qy	481	VNR\$M\$N\$L\$K\$R\$K\$T\$F\$A\$Y\$L\$A\$L\$A\$E\$P\$P\$K\$V\$G\$E\$F\$A\$K\$Y\$D\$L\$T\$T\$G\$V\$K\$K\$U\$Y\$G\$D\$N\$R\$Y\$G\$E\$P\$L\$P\$G\$E\$G	540
Db	481	vn\$r\$m\$n\$l\$k\$r\$k\$t\$f\$a\$y\$l\$a\$l\$a\$e\$p\$pk\$vg\$e\$fa\$ky\$d\$lt\$t\$g\$vk\$ku\$y\$g\$d\$nr\$y\$g\$e\$pl\$pg\$e\$g	540
Qy	541	ED\$E\$G\$I\$I\$C\$F\$V\$H\$D\$E\$T\$M\$K\$S\$E\$L\$Q\$I\$V\$A\$N\$V\$S\$E\$V\$E\$A\$Y\$V\$K\$L\$P\$R\$V\$Y\$G\$F\$H\$G\$T\$F\$G\$A\$D\$L\$A\$Q\$V	599
Db	541	ed\$e\$g\$i\$i\$c\$f\$v\$h\$d\$e\$t\$m\$k\$s\$e\$l\$q\$i\$v\$a\$n\$v\$s\$e\$v\$e\$a\$y\$vk\$lp\$r\$v\$y\$g\$f\$hg\$t\$fg\$a\$d\$la\$q\$v	599

RESULT	2
AAE04789	
ID	AAE04789 standard: Protein: 605 AA.

XX	AAE04789;	
AC		
XX		
DT	10-SEP-2001 (first entry)	
XX		
XX		
DE	Lycopersicon esculentum neoxanthin cleavage enzyme, LENCED1.	
XX		
KM	Tomato; neoxanthin cleavage enzyme; LENCED1; abscisic acid; ABA;	
KW	stress tolerance; transgenic plant; plant breeding; antisense-therapy;	
KM	plant growth protectant; herbicide.	
XX		
OS	Lycopersicon esculentum.	
XX		
PN	EP116794-A2.	
XX		
PD	18-JUL-2001.	
XX		
PF	11-JAN-2001; 2001EP-0300218.	
XX		
PR	13-JAN-2000; 2000JP-0010056.	
XX		
PR	11-JAN-2001; 2001JP-0003476.	
XX		
PA	(RIKE) RIKEN KK.	
XX		
PI	Iuchi S, Kobayashi M, Shinozaki K;	
XX		
DR	WPI: 2001-400081/43.	
XX		
DR	N-PSDB; AAD09401.	
XX		
PT	A DNA encoding a protein with a neoxanthin cleavage activity for	
PT	producing transgenic plants with improved or decreased stress tolerance	
XX		
PS	Claim 3; Fig 2; 101pp; English.	
XX		
CC	The invention relates to neoxanthin cleavage enzymes and their	
CC	corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key	
CC	role in endogenous abscisic acid (ABA) biosynthesis under drought stress.	
CC	Neoxanthin cleavage enzyme is used for improving stress tolerance in a	
CC	plant when expressed in a plant cell. The invention also relates to	
CC	methods for increasing or decreasing stress tolerance in a plant by	
CC	introducing the DNA into the plant, and a transgenic plant into which a	
CC	neoxanthin cleavage enzyme is introduced. The improvement of stress	
CC	tolerance in plants is useful, for example in plant breeding. Neoxanthin	
CC	cleavage enzyme genes are useful for producing transgenic plants. An arid	
CC	land can be improved by growing transformant weed for several years and	
CC	then removing the weed by specifically lowering stress tolerance in the	
CC	weed by inducing an inducible promoter. The present sequence is	
CC	Lycopersicon esculentum neoxanthin cleavage enzyme, LENCED1 protein	
CC	related to the invention.	
XX		
Sequence	605 AA:	

Query Match	72.48;	Score 2280.5;	DB 22;	Length 605;
Best Local Similarity	71.78;	Pred. No. 4.4e-222;		
Matches 435; Conservative	63;	Mismatches 90;	Indels 19;	Gaps 6;

OY	5	TATAAASORBLGNGHTOPRPLSSOSSOLDSTYS---	SILPMASRYTRKLNWSSALHTTPALH	61
		: :	: :	
Dd	3	tltasatctwtl----	kkltsmpskceftgfatnsatsltlnqhnrgcsnlntssldqaplhl	57
		: :	: :	
OY	62	PFKOSSN--SPA--TVVPPKAKESN----	*KONMLFORAAAALLDAEGLVLSHEKLHP	112
		: :	: :	
Dd	58	fpgsgsnypqrknmtisipbkqgnnnsssstskwnlvqkaamaIdavesal	tkhelehp	117
		: :	: :	
OY	113	LPRKTADPSVOIAGNEFAPNEQDPVRNRNLPPVVGKLEDSIKGVYVNRGANGAPLIEPTGTGHFFD		172
		: :	: :	
Dd	118	lpktadprvqtisnftaprvpenrvogqsilpvgtkllpkcvggyvvynrganrl	leptcaighffid	177
		: :	: :	
OY	173	GDOGWHAVKFEHGSAASYACRTQTGNREYQEROLGRPVYPKAIGELHGHTGIARLMIFYAR		232
		: :	: :	
Dd	178	gdgmwhavqfngksassycrtteerlvvkekalrpyrfpkalgblshslarlmifyar		237
		: :	: :	

[illegible]

CC		resistance of a plant to a pathogen such as fungus, virus, bacterium,
CC		nematode or insect (e.g. European corn bore), preferably
CC		Sclerotinia spp., Phoma spp., or Phomopsis spp., by stably incorporating a
CC		construct containing the gene into the genome of the plant. The gene is
CC		useful for regulating gene expression in a plant, in response to a
CC		stimulus such as infection with a pathogen, damage from a pathogen,
CC		hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,
CC		oxalic acid or expression of a gene encoding oxallic acid oxidase. The
CC		genes are also useful for stem-preferred regulation of gene expression in
CC		a plant. The genes are useful in agriculture, particularly in the
CC		breeding of crop plants with improved agronomic traits, for modifying
CC		abscisic acid (ABA) metabolism and for modifying amino acid transport and
CC		content in plants. The present sequence represents a neoxanthin cleavage
CC		enzyme-like protein from Lycopersicon esculentum used in the
CC		characterisation of sunflower NCE.
SO	Sequence	605 AA:
QY	Query Match	72.4%; Score 2280.5; DB 22; Length 605;
Db	Best Local Similarity	71.7%; Pred. No. 4,4e-222;
Matches	Conservative	63; Mismatches 90; Indels 19; Gaps 6;
QY	5 TATAATGCGWLGNGNHTQPPLSSOSSDLSYCS--- <td>61</td>	61
Db	3 ttttatatltwtl-----kklmpsskefgfaansislknqhmrgslnnsldgppllh	57
QY	62 FKKOSNN--SPA--IYVKRKAKESN-----TGMOMFORAAAALDAAGFLVSHKLRP	112
Db	58 ftkpsnygtprkcnnctishpkqgemnnssstskwnlvvgdaaaladavesalkthelehp	117
QY	113 LPTADPSVOLAGNEFAPVNEDPPVRNRNLPVVGKLPDSIKGVYVRNGANPLRHEPYTGHPRD	172
Db	118 lptktadvrqtgisgnfayrpempvcsglrpfytlrkpcvqvgyuzryngnprifteraghhftd	177
QY	173 GGCMWNAAKFEEHGSAZYACRFQTOTNFVOEROLGRVFKAIGELNHGTARLMFYAR	232
Db	178 gggmwhavvfkfnsgasyacrffeterlyvekalgrpvfpkaigelhghgiarlmfyar	237
QY	233 AAGAVIDPRAHGGVANAGLIVYFNGLRIAMSEDDLPROYOITPNRGDLTKVTGRFDQLRS	292
Db	238 glflglvdhsikgvgvaagaqlvytfmrillamseddiprvnvkrtpcgdlktegrtfdidgqlks	297
QY	293 TMIAHPKYDVESGETLFALSIDYVASKEPLYKFRFPSPDGTSPPVELOLDPTMWHDFATE	352
Db	298 tmiahpkiidvpsgeellfalsiydvrlqkyrlkfytfskgekndevpredtmhdteate	357
QY	353 NFVVVPDOOVFKLPBMIRFGSGPVIYDKKKVARFGLIDKYAEDSSNIKITADPCFHLL	412
Db	358 nfvvpdpdgqvvrfxmsemlirgspvvdydkkvserfglldkyakdsqdlkwverpdccfcfhll	417
QY	413 NNMAEPEPEDVAVYGSCMTPPDISFTNESDENILSVLSIRLNLTKGESTRRPIISNMO	472
Db	418 nnaweeaeedelvvisgsmtcpdisinecdeglkvsylseitrlnlkkyskctrsfsltenpe	477
QY	473 QYNLEAGMNRMMLGRKTFAFLATALAEPPMKYSGFEAKVDLTGTGEVKNHYGDNRKYGEBL	532
Db	478 qynleagmnrmnlkgktcfeayylaalaepmkysgfekvnllfgevkekflygdnkkyggerl	537
QY	533 FLIPGF--GGEDEBGTYLCFVHDNBKTMSELQIVNAVASLEVETAVKLSPRVPGHFSTFG	590
Db	538 flipdrnskreedddyllaetvhdekewkselqivnavmslkleatvkipsrvpygbtgftfin	597
QY	591 ADDLIARK	597
Db	598 anlclang	604
RESULT	4	
ID	AAE04787	
XX	standard: Protein; 612 AA.	
AC	AAE04787;	

XX 10-SEP-2001 (first entry)
DT
XX
DE Vigna unguiculata neoxanthin cleavage enzyme, CPRD65.
XX
XX Cowpea: neoxanthin cleavage enzyme; abscisic acid; ABA; herbicide;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant; Cowpea Responsive to Dehydration; CPRD65.
XX
OS Vigna unguiculata.
XX
PN EP1116794-A2.
XX
PD 18-JUL-2001.
XX
XX 11-JAN-2001; 2001EP-0300218.
XX
PF 13-JAN-2000; 2000JP-0010056.
PR 11-JAN-2001; 2001JP-0003476.
XX
XX (RIKE) RIKEN KK.
PA
XX
XX Iuchi S, Kobayashi M, Shinozaki K;
PI WPI: 2001-400081/43.
XX
DR N-PSDB: AAD093399.
XX
XX A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance
XX
XX
PS Claim 3; Fig 2; 101pp; English.
XX
XX The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present sequence is
CC Vigna unguiculata neoxanthin cleavage enzyme, CPRD65 (Cowpea responsive
CC to dehydration) protein. CPRD65 gene is isolated from cowpea plant.
XX
XX Sequence 612 AA;
XX

	Query Match	68.8%	Score 2168.5;	DB 22;	Length 612;
	Best Local Similarity	68.6%;	Pred. No. 1.le-210;		
	Matches 423; Conservative	61;	Mismatches 98;	Indels 35;	Gaps 8;
Oy	8 AAVSGRWLGAGNTGTOPPLSSOSSSDLSYCSLSPMAARVTRKLN-VSSALHTPALHPKO-	65			
Dd	3 ssaentfnatlprrpfkdlpsr--spenlrl--rksssnlticsqlt---lhfpqa	56			
Oy	66 ----SSNSPAlYVRKAESNT-----KOMNLFQRAAAALDAA	100			
Dd	57 gptststactttcprclktitttttcttppretncplsdtnqplrpqawntfqlkaaalaldv	116			
Oy	101 EGFVTSHEKLRPLPTADPSVOIAGNFAPVNEQPVRRLPVVGKLPDISIKGVYENGANP	160			
Dd	117 etalvshekrhlprlptadprvgjagnfapryphnaadgglpvngklpkcdldgyuvynanp	176			
Oy	161 LHERPTGNHFRGDGMVAHVAKFEFBSASYSACSFOTQNRNVDRQIGRPFFRAIGELSH	220			
Dd	177 llyepaghnhffidogdmvahnvklftngaasactftqtisrqslgtrfbraklgelghn	236			
Oy	221 TGIALMLTFVAAAAGIDVPAGHGTVANAAGLVYFNGLRIAMSEDDLPVOVQTTPMGDKLT	280			

Db	237	sgiaelllffgarqlfglvdsqgmgyanaaglyfnnhlllamseddlyphvritpngdlitc	296
Qy	281	VGREDFGQLESTWIAAPKVDPESEGLFALSYDVVSKPYLKTFRSPDGTNSPDVEIOLD	340
Db	297	vgrydfngqlnsttmahprklidrvgdalhatsydvilqkpylkyfrfsapdvkspdvlelplk	356
Qy	341	QPTMMHDAITENFVVVPRDOOVYFKLPFEMIRGSSPVYVKNKNVARGILDKAEDSSNIK	400
Db	357	epctmhdalientlvvvrdqvvfkltlemtlgssprvydknktstfgljlnknakdanamr	416
Qy	401	WIDAPDCFCFLTMAMEEPETDEVVVVGSCMPRPDSIFNPESENLKSVLEIRLMLKGE	460
Db	417	wldapdcfcfltmameepeteelvvvgscmtrpdsifneceslksvlselrlhrlrgk	476
Qy	461	STRPRTISENOQVNLNAGWVNRNMLGRKTKRKYLLALAEPMKPVGSFAKVDLTGSEVKN	520
Db	477	strprtlis-daeygnvlaagmwvnrnlgrtkrqfaylalaepwkvsgfakvdlisgevvky	535
Qy	521	LYGDNRVYGEPLFPRGSGGEDEGYLTCFVHDEKVTWKSLEIOTVMAVSLVEATVTLPSRV	580
Db	536	mygeekfyggeplflpr-ngkqeddyllatvndekewkselqlvnaqlkleasiklpsrv	594
Qy	581	PYGFHGTFIGADDLAKQ	597
Db	595	pygfhgtrfihskdlrkq	611
RESULT 5			
AAE04782	AAE04782 standard; Protein; 583 AA.		
XX	AC	AAE04782;	
XX	DX	10-SEP-2001 (first entry)	
XX	DE	Arabidopsis thaliana neoxanthin cleavage enzyme; AtNCD1.	
XX	KM	Neoxanthin cleavage enzyme; AtNCD1; abscisic acid; ABA; herbicide;	
XX	KW	stress tolerance; transgenic plant; plant breeding; antisense-therapy;	
XX	KW	plant growth protectant.	
OS	Arabidopsis thaliana.		
XX	PN	EP116794-A2.	
XX	PD	18-JUL-2001.	
XX	PE	11-JAN-2001; 2001EP-0300218.	
XX	PR	13-JAN-2000; 2000JP-0010056.	
XX	PR	11-JAN-2001; 2001JP-0003476.	
XX	PA	(RIKE) RIKEN KK.	
PI	Iuchi S, Kobayashi M, Shinozaki K;		
DR	WPI: 2001-400081/43.		
DR	N-PSDB; AAD09394.		
PT	A DNA encoding a protein with a neoxanthin cleavage activity for		
PT	producing transgenic plants with improved or decreased stress tolerance		
PT	-		
PS	Claim 3; Fig 10; 101pp; English.		
CC	The invention relates to neoxanthin cleavage enzymes and their		
CC	corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key		
CC	role in endogenous abscisic acid (ABA) biosynthesis under drought stress.		
CC	Neoxanthin cleavage enzyme is used for improving stress tolerance in a		
CC	plant when expressed in a plant cell. The invention also relates to		
CC	methods for increasing or decreasing stress tolerance in a plant by		
CC	introducing the DNA into the plant, and a transgenic plant into which a		

CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present sequence is
CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE1 protein.
CC The AtNCE1 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC library using a cDNA of the cPRD5 (Copeia Responsive to Dehydration)
CC gene isolated from copeia plant as a probe.

XX Sequence 583 AA;

Query Match 63.2%; Score 1991; DB 22; Length 583;

Best Local Similarity 64.5%; Pred. No. 1e-192; Indels 22; Gaps 6;

Matches 380; Conservative 82; Mismatches 105; Indels 22; Gaps 6;

QY 16 GGNHTQPLSSQSSDLSY--CSSLPMAASVYTRKLNVSALTPTPALHFPKSSNSPAIV 73
DB 11 gqiktwp----gaqldlfrpikrpkvktcvqdvre-lltkrqglftrptatpp--- 62

QY 74 VKPRAKESNTKQMLFQRAAAALDAAGFLVSHKHLPRKTDADPSVOIAGNPAVMEQ 133
DB 63 -----ghmrlrlnlfqkaaaataaeraalishedsplrpktadprvqiaqny3vpes 116

QY 134 PYRRNLPRVGLRPSDIKGVYVRNGANPLHEPVTGHFFDQGMVAVFEHGSASVACRF 193
DB 117 svrrnlvteglrpdcdldgvytrnganpmfpeaghhllfdgdmvnavhktngsasyacrif 176

QY 194 TQTNRFVQEROLGRPFVFKAIIGELHGTGIALRLMFLYARAAGIYDPAHGVANAGIVY 253
DB 177 tkterlvgeklrgrpvfkaigelhghsgiarlmfayargicqlinnqnygvanaaglyv 236

QY 254 FNGRLAASDDLPQVOYITPBGDLKTVGRPDFGQLSMTAHKVPDESGELFALSTD 313
DB 237 fmrllamseddlypykltitgvdltvgyrdfgqlksamiahpkldpvtkelhalsyd 296

QY 314 VYSKPYLKYFRFSPGTSKSPDVEIQLDQPTMMHDFAITENFVVPDQOVFKLPRMIMG 373
DB 237 vvkkylykfifspogvkspeleiprcmndfaiteenivpdpqgvvixlgemls9k 356

QY 374 SPVVYDKNKVARFGILDKVAEDSSNIKWIDAPDCFCFLHNMABEPTEDEVVIGSCMP 433
DB 357 spvvdgkcvkvarlgimpdateaagilwnspetcfhlmwavespceceivviscmap 416

QY 434 PDSINESDENIKSVLSIEIRLNLKTGSETRRPIISNEDQOVNLEAGVNRNMLGRKTKFA 493
DB 417 adslfnerdeslrsvlseirlnlrtrktrslivned--vnlleigmvnrnmlgrktrfa 474

QY 494 YLALAEPPKVSQFAKVLDTTGEVKKHLVGNRBYGGEPLFLPDEGG---EEDEGYIICF 549
DB 475 flalalyppkvsqfakvdlcvgemkkylyggekysg9epflfprgnsgeeneeddyilfch 534

QY 550 VHDETWKSLELOIVNAVSLVEAGTVKLSRPVYGFHGTGFIADLAKOV 598
DB 535 vhdetskselqilnavnlklatitklpsrpygfhgtfvsdnelvdql 583

RESULT 6
AAE04788
ID AAE04788 standard; Protein; 604 AA.

XX AAE04788;

XX 10-SEP-2001 (first entry)

XX Zea mays neoxanthin cleavage enzyme, VP14.

XX Maize; neoxanthin cleavage enzyme; VP14; abscisic acid; ABA; herbicide;
KM stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant.

OS Zea mays.

XX EPI116794-A2.

XX 18-JUL-2001.

XX 11-JAN-2001; 2001EP-0300218.

XX 13-JAN-2000; 2000JP-0010056.

XX 11-JAN-2001; 2001JP-0003476.

XX (RIKE) RIKEN KK.

XX Iuchi S, Kobayashi M, Shinozaki K;

XX WPI: 2001-400081/43.

XX N-PSDB: AAD09400.

XX A DNA encoding a protein with a neoxanthin cleavage activity for
XX producing transgenic plants with improved or decreased stress tolerance

XX Claim 3; Fig 2; 101pp; English.

CC The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present sequence is Zea mays
CC neoxanthin cleavage enzyme, VP14 protein related to the invention.

XX Sequence 604 AA;

Query Match 61.3%; Score 1930; DB 22; Length 604;

Best Local Similarity 63.3%; Pred. No. 1.7e-186;

Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;

QY 22 PPLSSQSSDLSYSSSLPMASVYTRKLNVSALHNP---PALHFPKSSNSPAIVV---- 74
DB 17 pararasas-----nsvtrspravassvpaeclqapfhhkpvadlpapsrkpaalavpgha 71

QY 75 -KPKAKESNTKQMLFQRAAAALDA-AGFLVS-HEKHLPRKTDADPSVOIAGNPAVYN 131
DB 72 aaprkaeagkkqnlifqraaaaaldaeegivanylerphgipstsdavqiaagfapag 131

QY 132 EQPVRRNLPRVGLRPSDIKGVYVRNGANPLHEPVTGHFFDQGMVAVFEHGSASVACRF 190
DB 132 erpyvhelprvsgtrppfidgvyarrnganpccfprvgghnldfgdgmwhalrlrngaeesya 191

QY 191 CRTQTNRVQEROLGRPFVFKAIIGELHGTGIALRLMFLYARAAGIYDPAHGVANAG 250
DB 192 crftetarlrgeralgrpvfkaigelhghsgiarlmfayaraagilvdpsasgltvanaag 251

QY 251 LVYFNGRLAASDDLPQVOYITPBGDLKTVGRPDFGQLSMTAHKVPDESGELFAL 310
DB 252 lvyfngrlamseddlypyhvraaddgdlvgyrdfgqlgcamaiahpkldpatgelhal 311

QY 311 SYDVSKPYLKYFRFSPGTSKSPDVEIQLDQPTMMHDFAITENFVVPDQOVFKLPRMI 370
DB 312 sydvskipylykfiftrpogtksdaveiplegpctmndfaiteenivvpdhqyvfalgeml 371

QY 371 RGSFPVYDKNKVARFGILDKVAEDSSNIKWIDAPDCFCFLHNMABEPTEDEVVIGSC 430
DB 372 rgspvrvldkcktsrfyglpkhaadasemawdvdpdcfcfhlmwadesatgevvnvlgsc 431

Oy	431	MPPPOSIFNESEDNKSVSEIRPLNKTGSETRPIISNDQOVNLGAGVNNMLGRKT	430
		: : : : :	
Dd	432	mfpadsifnesderlesviteirldarigrstrravlp-psqgenlevgnvnrllgres	430
Oy	491	KFAYLALAPMPKPVSGFAVDLTGTGVKKHLYDNNKGEGEPFLPGPGGE-----EDEGX	545
		: : : : : :	
Db	491	rYayLaeeapwkesgfakedisgtgalktfeygegrfgepcrcvpmdaahprgeddy	550
Oy	546	IICFVHADETKMSEIQIVNAVSLVEAVTYKLPSRVYPGFHGTEIGADILAKQ	597
		: : : : : : : :	
Dd	551	vltfvnderagsellivnaadirlatvqlpsrvpfvgftgltgtleaq	602
 RESULT 7			
ID	AAB72309		
xx	AAB72309 standard; Protein; 604 AA.		
AC	AAB72309;		
xx			
DT	16-MAY-2001 (first entry)		
xx			
DE	Neoxanthin cleavage enzyme-like protein amino acid sequence.		
xx			
KW	NCE: amino acid permease; AAP; glutamic acid rich protein; GRP;		
KM	pathogen resistance; abscisic acid metabolism.		
xx			
OS	Zea mays.		
PN	WO200112801-A2.		
PD	22-FEB-2001.		
xx			
PF	17-AUG-2000; 2000MO-US22961.		
PR	18-AUG-1999; 99US-0149656.		
PR	23-MAY-2000; 2000US-0206405.		
PA	(PION-) PIONEER HI-BRED INT INC.		
PA	(CURA-) CURAGEN CORP.		
PI	Bidney DL, Crasta OR, Hu X, Lu G;		
DR	WPI: 2001-211215/21.		
PT	Novel isolated defence-related signalling gene isolated from sunflower		
PT	encoding neoxanthin cleavage enzyme, amino acid permease or glutamic		
PT	acid-rich protein useful for increasing resistance of plant to a		
PT	pathogen -		
xx			
PS	Disclosure: Fig 1; 135pp: English.		
xx			
CC	This invention relates to defence-related signalling genes isolated from		
CC	the sunflower (<i>Helianthus annuus</i>). The genes encode a neoxanthin cleavage		
CC	enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich		
CC	protein (GRP). The signalling gene is useful for increasing the		
CC	resistance of a plant to a pathogen such as fungus, virus, bacterium,		
CC	nematode or insect (e.g. European corn borer), preferably		
CC	<i>Sclerotinia spp.</i> , <i>Phoma spp.</i> , or <i>Phomopsis spp.</i> , by stably incorporating a		
CC	construct containing the gene into the genome of the plant. The gene is		
CC	useful for regulating gene expression in a plant, in response to a		
CC	stimulus such as infection with a pathogen, damage from a pathogen,		
CC	hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,		
CC	oxalic acid or expression of a gene encoding oxalic acid oxidase. The		
CC	genes are also useful for stem-preferred regulation of gene expression in		
CC	a plant. The genes are useful in agriculture, particularly in the		
CC	breeding of crop plants with improved agronomic traits, for modifying		
CC	abscisic acid (ABA) metabolism and for modifying amino acid transport and		
CC	content in plants. The present sequence represents a neoxanthin cleavage		
CC	enzyme-like protein from Zea mays used in the characterisation of		
CC	sunflower NCE.		
xx			

Seq	Sequence	604 AA:
Query Match	61.3%; Score 1930; DB 22; Length 604;	
Best Local Similarity	63.3%; Pred. No. 1.7e-16;	
Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;		
QY	22 PRLSSQSSDLSYCSLSLPMASNVTKKLNVSSALHP-----PALHFPKQSSNSPAIVY-----	74
DB	17 parrasars-----nsvfsprravssvpaeclqgrflfkpradlpaprsrkpaalavpgha	71
QY	75 -KPKAKESNTKOMNLFORAAALDA-KEGFLVS-HEKILHPKPADSDSVQIAGNFAPVN	131
DB	72 aapraegggkqlnlfqraaaaaldaeegfvaanvlerphnlpstadaqvlgagfayvg	131
QY	132 EOPVRNRLPVNGKLPDSTIKGYVVRNGANPLDEBPVTGHHFFEDGDCNVHAVKREHGSA-SYA	190
DB	133 erprvhelprvsgripfrfidgyatrganarpcfrvaghnlfdqgdmvhalirngaaesya	191
QY	191 GREFOTNPFVOROLGRPFVPRKATIGELNGHNGIARLMLFYRAAAGYIDPAHGTGVANAG	250
DB	192 cftccatrlrgeraligrvfrpaalgetlnghsialralfyraaagylvpsagfgyanaq	251
QY	251 LVYENGRLAMSEDDLPYQVOTTPNGDLKTVGRFDFDQLESTIMAHKVPESGELPAL	310
DB	252 lvyfngirlamseedlpyhvravadgdletvgyrfdqglgcamiahpkldpatgelhal	311
QY	311 SYDVVSKRYLTKFRKSPDGTSPRYEITOLDQPTMMHDEATENFVVVVDQVVKRLPMI	370
DB	312 sydvtkrpylkyftrpqrqksdvdvlebpctmhdfatefenvrvrdhqvklqeml	371
QY	371 RGSFVVVVDKKNVAFGGLDKYAEDESSNIKWIDAPDCFCFHLNMAEBEPTEDEVVLTSC	430
DB	372 rggsrvvldktektrsfyglprkhaadasemawdyrdccfhlmawedatagevvvlgsc	431
QY	431 MTPPDSINESDENIKSVLSERLNLKGTGSTRPTISNEDQVNLKGMVNRNMLGRKT	490
DB	432 mpradslinesderlesvltelridatgrstlravlp-psqgenlvgmvrnllgtes	490
QY	491 KRAYALALBPMPKVSGFPAKVDLTGVEVKHLYGDMRYGSGEPLFLPREGGE-----EDGXY	545
DB	491 ryayalavepmpkesgfakedstgelctkfeyyeggrfgyepcfvpmrpaahprgeddy	550
QY	546 ILCFVHDEKTMKSELQIVNAVSLLEVATVKKLPSRYVYGFGHTFGADDLAKQ	597
DB	551 vltfvhderagyselllvnaadirlleaetvqfprsvrpfghgtftlgtgleaq	602
RESULT 8		
AAE04786	AAE04786 standard; Protein; 577 AA.	
XX	AAE04786;	
AC	AAE04786;	
XX		
DT	10-SEP-2001 (first entry)	
XX		
DE	Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5.	
XX		
KW	Neoxanthin cleavage enzyme; AtNCED5; abscisic acid; ABA; herbicide;	
KM	stress tolerance; transgenic plant; plant breeding; antisense-therapy;	
XX	plant growth protectant.	
OS	Arabidopsis thaliana.	
PN	EP1116794-A2.	
PD	18-JUL-2001.	
PF	11-JAN-2001; 2001EP-0300218.	
PR	13-JAN-2000; 2000JP-0010056.	
XX	11-JAN-2001; 2001JP-0003476.	

D	b	95	tfidp-----psrpsvopkhuvslnfnrpyvidelprptccellngtlprlsingaylrnmp	149
O	y	161	LHEPVtGHFFDGDGMVNAVKEHGSASVACRFtQTNRFEWEROGRPVEPKAIGELIGH	220
D	b	150	qflprgpyrlhgdgmhaikinhogkatlsryukytkyuvexqtagavmpnfsgfngy	209
O	y	221	T-GARLMLFVBARAAGVDPABHGTGVANNGLVYENRLLAMSBDLrPOVOtTPRNDK	279
D	b	210	tasvargalltaarylbtqynpvnigslanslafrnrltalaesdipravrillesgdle	269
O	y	280	TVGRFDEFGDLESTMJNAHPKVPDESGLFALSYDVNSKPYLKFRRSPDCKSPDEI-Q	338
D	b	270	tlgydfdgqslamstahpkrqrltgctafatrygrv-prfltyfridsagkkqrdwylfe	328
O	y	339	LDOPRNMHDAITENFVNVPRDOQVFR---LPRMIRGSGSVYVDKKNVAFGLIDKTAED	355
D	b	329	mtpsfldldatklrhaifaeilqgmnmmdlrvlegsspyvntngtlprlgyipkryagd	368
O	y	396	SSNtKWIDAPRCFCFHMNAEERPEtDEBVVIGSGCMPPDSIFENESDENKSVLSERLN	455
D	b	389	esemkwefevpfnllhalnawdeddgnsvllapnlnsichtlern-dlhalvekykid	447
O	y	456	LKTGESFRPRFISNEDDOVNLGAGVNRNMIGKRTKATYALALDEPWPVGSKAKVDITGS	515
D	b	448	lvtlvtrhpisad-----nldfavlhpaflycstosryvaaiqdmpmklsjgvvklavskg	502
O	y	516	E-----VKHhGYNRRYGEPELrFPGEGS---EEDGYtLlCFHNDKTKWSELOJNAV	566
D	b	503	drddctvaarrmgyggcyggygefffaardprgpeaeedgyvvtyvhdevteskflvmak	562
O	y	567	S--LEVEATVKLPSRVPGHGTIGADDLAK	596
D	b	563	speleivaavvlprlrvpygfnlgltvkesdlnk	594

RESULT 10
AAE04783
ID AAE04783 standard; Protein; 595 AA.
XX
XX
AC AAE04783;
XX
DT 10-SEP-2001 (first entry)
XX
DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE2.
XX
KW Neoxanthin cleavage enzyme; AtNCE2; abscisic acid; ABA; herbicide;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant.
XX
OS Arabidopsis thaliana.
XX
PN EP116794-A2.
XX
PD 18-JUL-2001.
XX
PE 11-JAN-2001; 2001EP-0300218.
XX
PR 13-JAN-2000; 2000JP-0010056.
PR 11-JAN-2001; 2001JP-0003476.
XX
PA (RIKE) RIKEN KK.
XX
PI Tsuchi S, Kobayashi M, Shinozaki K;
XX
DR WPI; 2001-400081/43.
DR N-PSDB; AAD09395.
XX
PT A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance
XX
XX Example 10; Fig 10; 101pp; English.
XX

CC The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present sequence is
CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2 protein.
CC The AtNCED2 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC library using a cDNA of the CPR65 (Cowpea Responsive to Derradation)
CC gene isolated from cowpea plant as a probe.
XX Sequence 595 AA;

Query Match	30.7%;	Score 968;	DB 22;	Length 595;
Best Local Similarity	36.7%;	Pred. No. 8.3e-89;		
Matches 210; Conservative	166;	Mismatches 220;		Indels 36; Gaps 13

[illegible]

RESULT	11
AAB72306	
ID	AAB72306 standard; Protein; 595 AA
XX	
AC	
XX	AAB72306;
XX	
DT	16-MAY-2001 (first entry)

XX	Neoxanthin cleavage enzyme-like protein amino acid sequence.
DE	
XX	
KW	Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;
KM	NCB; amino acid permease; AAP; glutamic acid rich protein; GRP;
KM	pathogen resistance; abscisic acid metabolism.
XX	
OS	Arabidopsis thaliana.
XX	
PN	WO200112801-A2.
XX	
PD	22-FEB-2001.
XX	
PF	17-AUG-2000; 2000WO-US22961.
XX	
PR	18-AUG-1999; 99US-0149656.
PR	23-MAY-2000; 2000US-0206405.
XX	
PA	(PLON-) PIONEER HI-BRED INT INC.
PA	(CURA-) CURAGEN CORP.
XX	
PI	Bidney DL, Crasta OR, Hu X, Lu G;
XX	
DR	WPI; 2001-211215/21.
XX	
PT	Novel isolated defence-related signalling gene isolated from sunflower
PT	encoding neoxanthin cleavage enzyme, amino acid permease or glutamic
PT	acid-rich protein useful for increasing resistance of plant to a
PT	pathogen -

Example 1; Fig 1; 135pp; English.

CC This invention relates to defence-related signalling genes isolated from
CC the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage
CC enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich
CC protein (GRP). The signalling gene is useful for increasing the
CC resistance of a plant to a pathogen such as fungus, virus, bacterium,
CC nematode or insect (e.g. European corn borer), preferably
CC *Sclerotinia* spp., *Phoma* spp., or *Phomopsis* spp., by stably incorporating a
CC construct containing the gene into the genome of the plant. The gene is
CC useful for regulating gene expression in a plant, in response to a
CC stimulus such as infection with a pathogen, damage from a pathogen,
CC hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,
CC oxalic acid or expression of a gene encoding oxalic acid oxidase. The
CC genes are also useful for stem-preferred regulation of gene expression in
CC a plant. The genes are useful in agriculture, particularly in the
CC breeding of crop plants with improved agronomic traits, for modifying
CC abscisic acid (ABA) metabolism and for modifying amino acid transport and
CC content in plants. The present sequence represents a neoxanthin cleavage
CC enzyme-like protein from *Arabidopsis thaliana*, used in the
CC characterisation of sunflower NCEs.

Query Match	30.78;	Score 968;	DB 22;	Length 595;
-------------	--------	------------	--------	-------------

Best Local Similarity 36.7%; Pred. No. 8, 3e-89;
Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13

42 SRVTEKLNVSALHTPPALHFPKROSSNPAYVVKPKAKESNTKÖMILFÖRÅAALDAE 101

```

Db      | : | : | | | : | : | | | |
42 spitpsdmndrnkpktlh-----nrtnthtlvspcklirpemtlatalf-----ttvedyin 94

```

QY 102 GFLVSHKHLPLKPTADPSVQTAGNFAPVNEQPYRRLPLVY-GKLEPDSIKGYVVRNGANP 160

```

Db      95  tffdp-----psrpsvdpkhvnsdfnafpVldelpptdcclhngtVlpIsIngayirngnp 149

```

161 LHEPYTGHHFEDGDMVHAKFEHGSASYACRFTQTNRFRVQEROLGRPVFPRAIGELHGH 220

```

db 150 qf1prgpyhlfcdgdm1haik1hngkat1csryvktkykynvektqgavmpnvnfsgfngv 209

```

221 T-GAARLMFLARAAGIVDPAGTGVANAGLVYENGRLAMSEDDIPYQVITPNCGLK 279

```

Db 210 tasvaargal taarvlitqgynpvinglantslaftsmrlfalgaeotalpyavilteesgdlie 269
Qy 280 TVGREDFDQGLESTMIAHPKVPDPESGELFALSYDVWSKPKLYFRFSPDGTSKSPDVEI -O 338
Db 270 tlgrydfdfgkrlamsmtanhpkdrptltgettafarypvr -pflityfrfatsagkkgdrvpilfe 328
Qy 339 LDQPMMDFAITENFNVVVPDQOVFK---LPEMIRGSGSPVYIDKKNVAFRGLIDKRAED 395
Db 339 mtspslhfafatkrthalfaeiqglgmnmmlidlyleogspvtdngktrptlgyrlpkryagd 388
Qy 396 SSMKMWIDAPRCFCHLMNAEPEETDETVYVYISGCMPPDISFNEPDENKLSVLSERLN 455
Db 389 esemkewfeyvgfnlthanaadeddgnsvalapnalsientlermd -lvhalvevwxid 447
Qy 456 LKTGSTRPPLTISNDDQVNLNACMVNRNMIGRTKFAVYALALAEPMWPKVSGFAPKVDLTG 515
Db 448 lvtglvrrhplisar-----nldfavnlpafigrstcryyaaigdpmpklsygvvklldvskg 502
Qy 516 E-----VKKHLGDNRYGGEPLFLPBGEG---EEDEGYLTCFVNDERTKWSSELOJYNAY 566
Db 503 drdctvarnmygsqcygsepfvaridpgrpaeaeedgyvvtvlyvhaevtgeskflymdak 562
Qy 567 S--LEVEATVKLPSRVYRGFRHTIGADDLAK 596
Db 563 speleivaavrlprvpygfnqltfsksdlmk 594

```

RESULT 12

AAAG31334 standard; protein: 517 AA.

XX
AC AAG31334;

XX	17-OCT-2000	(first entry)
DT		

XX	Arabidopsis thaliana protein fragment SEQ ID NO: 37613.
DE	

XX Protein identification; signal transduction pathway; metabolic pathway;
KW

KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 KW

XX
OS
yy Arabidopsis thaliana.

AA
PN
XY

EP1033405-A2

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.

PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.

PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.

PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.

PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 10-APR-1999; 99US-0129077.

PR	19-APR-1999;	99US-013007/.
PR	21-APR-1999;	99US-0130449.
PR	22-APR-1999;	99US-0130510.

PR	23-APR-1999;	9905-0130510
PR	23-APR-1999;	9905-0130891
PR	28-APR-1999;	9905-0131440

PR 26-APR-1999; 99US-0131445
PR 30-APR-1999; 99US-0132048
PR 30-APR-1999; 99US-0132407

FN	00-MAY-1999;	99US-0132484.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485

PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.

PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149176.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144321.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145066.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146387.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149366.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0161997.
PR 29-OCT-1999; 99US-0162142.

Query Match 29.8%; Score 939; DB 21; Length 538;
Best Local Similarity 37.0%; Pred. No. 6.2e-86;
Matches 207; Conservative 106; Mismatches 198; Indels 48; Gaps 13;

QY 64 KOSNSPAIVKPK-AKESNTKOMLFDRAAALDAAGFLVSHKHLPLPKTADPSVO 122
DB 4 kisdgslisvhpksksgskldllderlvklm-----hdasipjh-----y 47
QY 123 IAGNFAPV-NEQVRRNLPVVGKLPDSIKGVYVRGANGAPLHEPVGHHFFDDGDGVAHWK 181
DB 48 lsgnflaprlrdecpvkdlprvnhgfllpeclngeftvrvgpnkldavagynhfdgdmhgyr 107
QY 182 FEHGSASVACRPTQTNRFVQEROLGRPVPPKAIIGELHGTGIARLMLFYARAAGIVDPA 241
DB 108 lkdgkatlystryvktsrlkgekfifaakfmk-igdlkgffglmwriqqlrtklklidnt 166
QY 242 HGTGVANAGLVYFNGRLAMSEDDLRYOVOTPNNGDLKTVGRFDEGOLSTMIAPKVD 301
DB 167 yngntaantalvyhghklalgaedkpyvklvledgdlqrlgldkrlthsfahpkvd 226
QY 302 PESGEFALSTYDVSKPRYIKRFPSPDGTSPDVEIQLDQPTMMDFATTEFVYVPPDQ 361
DB 227 pvtgemftfygs-hcpvlltyrlvlskdgmhdpvdlitsepimmdhfateyalftmdr 285
QY 362 VVEFKLPEMIRGSGPYV-YVKNKVARFGIIDKYAEDSSNKKWIDAPDCFFHLMNAEPE 420
DB 286 mhfrpkemwkekkmlysfopklatfgyvpryakelmlrwlfnoclfifhnanawe--e 343
QY 421 TDEVVIGSCMTPD----SIFNESDENLKSVLSEIRLNLKGTSTRPRTISNEDQOVN 475
DB 344 edevvlltrclenpdlmwsgvkkeklengfnelymfrfmktgsasqkksasa----- 398
QY 476 LEAGVNNRMKGRKTKFALALAEPPKYSGFAXKD-----LTTGVKKHLY-- 522
DB 399 vdfprlnecytgkkgqrvyvgtilldslakvtglikltdlhaeaetgrmlvevgnllgyldl 458
QY 523 GDNRYGGEPLFLPGEAGEDEGYIILCFVHDEKTMKSEIQIVNAVSLVEV--ATVKLPSRV 580

Db 459 gggrygsealyprcraeeddgyllffvndentgkscvclvldaktnsaepvavvelphrv 518
QY 581 PYGFHCTFGADDLAKQV 599
Db 519 pygfhlvtveeqlqetl 537

RESULT 14

AAAG13115
ID AAAG13115 standard; Protein; 539 AA.

XX
XX
AC AAAG13115;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 12488.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX
XX
PN EPI033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 990S-0121825.

PR 05-MAR-1999; 990S-0123180.

PR 09-MAR-1999; 990S-0123548.

PR 23-MAR-1999; 990S-0125788.

PR 25-MAR-1999; 990S-0126264.

PR 29-MAR-1999; 990S-0126785.

PR 01-APR-1999; 990S-0127462.

PR 06-APR-1999; 990S-0128234.

PR 08-APR-1999; 990S-0128714.

PR 16-APR-1999; 990S-0129845.

PR 19-APR-1999; 990S-0130077.

PR 21-APR-1999; 990S-0130449.

PR 23-APR-1999; 990S-0130510.

PR 28-APR-1999; 990S-0130891.

PR 30-APR-1999; 990S-0131449.

PR 30-APR-1999; 990S-0132048.

PR 04-MAY-1999; 990S-0132407.

PR 05-MAY-1999; 990S-0132484.

PR 06-MAY-1999; 990S-0132485.

PR 06-MAY-1999; 990S-0132486.

PR 07-MAY-1999; 990S-0132487.

PR 11-MAY-1999; 990S-0132487.

PR 14-MAY-1999; 990S-0132487.

PR 14-MAY-1999; 990S-0132487.

PR 16-JUN-1999; 990S-0139452.

PR 16-JUN-1999; 990S-0139453.

PR 17-JUN-1999; 990S-0139492.

PR 18-JUN-1999; 990S-0139454.

PR 18-JUN-1999; 990S-0139455.

PR 18-JUN-1999; 990S-0139456.

PR 18-JUN-1999; 990S-0139457.

PR 18-JUN-1999; 990S-0139458.

PR 18-JUN-1999; 990S-0139459.

PR 18-JUN-1999; 990S-0139460.

PR 18-JUN-1999; 990S-0139461.

PR 18-JUN-1999; 990S-0139462.

PR 18-JUN-1999; 990S-0139463.

PR 18-JUN-1999; 990S-0139464.

PR 18-JUN-1999; 990S-0139465.

PR 18-JUN-1999; 990S-0139466.

PR 18-JUN-1999; 990S-0139467.

PR 18-JUN-1999; 990S-0139468.

PR 18-JUN-1999; 990S-0139469.

PR 18-JUN-1999; 990S-0139470.

PR 18-JUN-1999; 990S-0139471.

PR 18-JUN-1999; 990S-0139472.

PR 18-JUN-1999; 990S-0139473.

PR 18-JUN-1999; 990S-0139474.

PR 18-JUN-1999; 990S-0139475.

PR 18-JUN-1999; 990S-0139476.

PR 18-JUN-1999; 990S-0139477.

PR 18-JUN-1999; 990S-0139478.

PR 18-JUN-1999; 990S-0139479.

PR 18-JUN-1999; 990S-0139480.

PR 18-JUN-1999; 990S-0139481.

PR 18-JUN-1999; 990S-0139482.

PR 18-JUN-1999; 990S-0139483.

PR 18-JUN-1999; 990S-0139484.

PR 18-JUN-1999; 990S-0139485.

PR 18-JUN-1999; 990S-0139486.

PR 18-JUN-1999; 990S-0139487.

PS Example 10; Fig 10; 101bp; English.

XX
CC The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present sequence is
CC Arabidopsis thaliana neoxanthin cleavage enzyme. AtNCED4 protein.
CC The AtNCED4 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC library using a cDNA of the CPD65 (Cowpea Responsive to Dehydration)
CC gene isolated from cowpea plant as a probe.

SQ Sequence 538 AA;

Query Match 29.8%; Score 938; DB 22; Length 538;
Best Local Similarity 37.0%; Pred. No. 7.8e-86;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;

OY 64 KQSSNSPAIVVPRK-AKESNTQOMNLFQRAAAALDAAGFLVSHKMLPDKTADPSVQ 122
DB 4 Kisdgsslsivnhrpsksgfskldlrlvkvlm-----hdasjplh-----y 47
OY 123 IAGNFAPV-NEQPVRRNLVVGKLRPSIKGVYVRNGANPLHEPVGTGNHFFDGDGVNAVAK 181
DB 48 Isgnfapirdeprkdlprvghflpeclngefvrvgpnokfdavagyhwhfdgdmhgyr 107
OY 182 FEHGSASVACRTQTRRFQVQROLGRPVPRKAIIGELHGTGIRLMLFYARAAGTIVDPA 241
DB 108 Ikdgkatyvsryvktsrlkqeeffgaaklmk-igdlkgffgllmvyvqglrtklkildnt 166
OY 242 HGTGVANAGLVFENGRLAMSEDDLPYOQIPIRNGDLTKVGRFPDGOESTMIAPKVD 301
DB 167 yngntaantalyhngklalgaedkpyvikvleogdlqclgildydkrlthstftanpkvd 226
OY 302 PESGELFALSYDVVSKNPKYLFKRFSPDGTSPDVEIQDQPTMMDFAITENFVVVPPDOQ 361
DB 227 pvtgemfifgys-heppyltyrviskdgmhdpvptlseplmmdfateteyalfmdlp 285
OY 362 VYFKLPEMIRGSGSPVY-IDKNKVAREGILDKYAEDSSNITKWDAPDCFCFHLNMAEPE 420
DB 286 mhfrpkemwkekmlysfdpkkarfgypryakdelmlrwlfeipncfifhnanawe--e 343
OY 421 TDEVVVIGSCMTPPD-----SIFNSEDENLKSIVLEIRLNLTKGSTRRPITISNEDQVN 475
DB 344 edevvlltcrlempdlamvsgkvkelenfneymfrnmktysasgkklisasa----- 398
OY 476 LEAGVNRNMLGRKTKFAYIALAEPPKVSQFPAKYD-----LTTGEVKKHLY-- 522
DB 399 vdfprlnecytgkkqryvgvtlidsiaktgltfdlhaeetgkrmlevgnikgylgl 458
OY 523 GDNRTGGEPLFLPREGGEDEGTIILCFVHDEKTKWSELOIVNAVSLVEE--ATVKLPSRY 580
DB 459 gegryysaealyvpretaeedgylifivndentgkscvltidaklmsaepvavveiphrv 518
OY 581 PYGFHGTFIGADDLAKOVY 599
DB 519 pygfhalvtveeqldgeqtl 537

Search completed: July 19, 2002, 10:11:42
Job time: 45 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2002, 10:11:03 ; Search time 21.04 Seconds
(without alignments)
2735.623 Million cell updates/sec

Title: US-09-758-269-6
Perfect score: 3150
Sequence: 1 MASFTATAVSGRWLGNNHT.....VPYGFHGFICADLIKQVY 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pirl:.*
2: pirl:.*
3: pirl:.*
4: pirl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2289	72.4	604	2 T51936	probable 9-cis-epo
2	2280.5	72.4	605	2 T07123	nine-cis-epoxycaro
3	2170.5	68.9	657	2 E96812	protein F3F9.10 [1
4	2044.5	64.9	589	2 A86425	probable 9-cis-epo
5	1991	63.2	583	2 T04531	nine-cis-epoxycaro
6	1930	61.3	604	2 T04351	viviparous-14 prot
7	968	30.7	595	2 T04438	hypothetical prote
8	938	29.8	538	2 T49193	neoxanthin cleavag
9	937	29.7	538	2 T51734	neoxanthin cleavag
10	745.5	23.7	475	2 AG1944	hypothetical prote
11	745	23.7	446	2 T17019	hypothetical prote
12	528.5	16.8	483	2 E87345	conserved hypotet
13	466	14.8	501	2 A70534	hypothetical prote
14	454.5	14.4	502	2 A70582	hypothetical prote
15	394	12.5	480	2 S76169	hypothetical prote
16	355.5	11.3	497	2 AE2341	hypothetical prote
17	352	11.2	484	2 JN0595	lignostilbene alph
18	317	10.1	489	2 S76206	lignostilbene alph
19	303.5	9.6	480	2 JCA324	lignostilbene alph
20	253	8.0	616	2 T10688	hypothetical prote
21	244	7.7	618	2 C84885	hypothetical prote
22	242	7.7	472	2 AG2417	hypothetical prote
23	238.5	7.6	556	2 F88115	protein F53C3.12 [
24	231	7.3	483	2 D87290	conserved hypotet
25	215	6.8	533	2 A47143	retinal pigment m
26	126	4.0	790	2 T34293	hypothetical prote
27	116.5	3.7	1832	2 T31113	mucin-like glycopr
28	114	3.6	870	2 S74291	hypothetical prote
29	113	3.6	466	1 IOEBV	replication initia

30	110.5	3.5	658	2 S60170	protein kinase Pak
31	108	3.4	437	2 G70019	conserved hypotet
32	107.5	3.4	658	2 T39500	serine/threonine-s
33	107	3.4	1461	2 T13157	mitotic checkpoint
34	107	3.4	26926	1 T38344	titlin, cardiac mus
35	106.5	3.4	1299	2 T47182	hypothetical prote
36	106	3.4	1679	2 T15968	hypothetical prote
37	106	3.4	3562	2 A47171	chondrotin sulfat
38	105.5	3.3	2591	2 T30288	pristinamycin I sy
39	105	3.3	2504	1 A57788	fatty-acid synthas
40	104.5	3.3	623	2 T52293	MYC-related DNA bi
41	104.5	3.3	1080	2 A35088	phycobilisome link
42	103.5	3.3	394	2 AG3616	hypothetical prote
43	103	3.3	775	2 B70449	phenylalanine--trN
44	103	3.3	4845	2 T31067	BTR repeat contain
45	102.5	3.3	531	2 AH3378	cell wall degradat

ALIGNMENTS

Query Match	Best Local Similarity	72.7%;	Score 2289;	DB 2;	Length 604;
Matches 432;	Conservative 71.3%;	Pred. No. 1.2e-165;	Mismatches 90;	Indels 18;	Gaps 5;
QY 5	TATAVSGRWLGNNHTPP	SSQS	SLSTYCS---	SLPMASRVTRKLVSSALHPRPALH	61
DB 3	TTTSHATNTWT----	KKPL	SMPSKEFGFASNS	ISLKNQHNRSLSLAINSLAPPLH	57
QY 62	FPKOSN-----	SPAIVV	KPAKESNT-----	KOMLFORAAALDAAGFLVSHKHLPL	113
DB 58	FPKOSNQTQRTKSTIS	HPKQ	ENNNSSSISKWNLVQRAA	AMALDAVGAALTKHEHLPL	117
QY 114	PKTADPSVQIAGNFAP	NEQV	RRNLPVYGLKLP	PSIKSGVYVRNGANPLHEPVYGHHPFDG	173
DB 118	PKTADPVRVQISGNFAP	VENP	VCOSLPVTKIPKQVQ	GVYVRNGANPLFEPTAGRHHPFDG	177
QY 174	DGMVAHVAFFEHGSAS	YACR	FTQTRNFVQEROLG	RVPRFKALIGELHGHGIALRMFLYARA	233
DB 178	DGMVAHVAQFKNGSAS	YACR	FTETERFVEKALGR	VFPRFKALIGELHGHGIALRMFLYARG	237
QY 234	AAGIVDPAHGTGCVAN	AGLV	YFNGRLAMSEDDLP	POVOTITNGDKLTGVRPFDGQLSEST	293
DB 238	LFGLIDHSRGTCGVAN	AGLV	YFNNRLAMSEDDLP	RVHVAVTPTGDKLTGGRDFDQGLKST	297
QY 294	MIAPKVPDESSEGE	LFALSYDVVSKPYL	KYFRSPDGTSPVEI	QLDQPTMMHDFATEN	353
DB 298	MIAPKLDVPVSGEL	FLFALSYDVYDIQ	KYLYFRFSKNGEKS	NDVEIPEVDPJTMHDFATIEK	357
QY 354	FVVVPDQGVVFKLE	PMIR	IGSGSPVYVDKKNV	ARFGILDKYADSSNIKKIDAPDCEFFILW	413
DB 358	FVILPDQGVVFKMS	EMIR	IGSGSPVYVDKKNV	ARFGILDKYADSSNIKKIDAPDCEFFILW	417
QY 414	NAMEPETFDEVV	IGSC	TPDSTIFNEDENL	KSVLSERINLKTG	ESTRRPITISNEDQ 473

Db	418	NAMEPELDELIVIGSCMTPPDSIFNDCBGLKSVLSELRNLKTKSTKRIIENPBQ	477
Qy	474	VNLEAGVNRNMIGRKTKEAYLALAEPPKVSQFAVDLTGGEVKKHLYGDNRYGEPLE	533
Db	478	VNLEAGVNRNRKLGKRKTQYAYLAIAPWPKVSGFAVDLTGGEVEKFIYGDNRKYGEPLE	537
Qy	534	LPGE-IGSEDEGYILCFVHDEKTMKSELQIVAAVSLVEATVYKILPSRPVYGFHGFTGA	591
Db	538	LPRPNKSEDEDDGYLLAFVHDEKEMSELDQIVAAMLKLEATYKILPSRPVYGFHGFTGA	597
Qy	592	DLAKQ	597
Db	598	NDLANQ	603

```

RESULT      2
T07123
nine-cis-epoxycarotenoid dioxygenase - tomato
N:Alternate names: Probable neoxanthin cleavage enzyme
C:Species: Lycopersicon esculentum (tomato)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C:Accession: T07123
R:Burridge, A.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z15934
A:Accession: T07123
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-605 <BUR>
A:Cross-references: EMBL:Z97215; PIDN:CA810168.1

```

Query Match	72.4%;	Score 2280.5;	DB 2;	Length 605;
Best Local Similarity	71.7%;	Pred. NO. 5.1e-165;		
Matches 435; Conservative	63;	Mismatches 90;	Indels 19;	Gaps 6;

Oy	5	TATAAVSGRMJGCHHTOPPLSSOSSDLSICS---	SLPMAKVTYTKRLVNSALMLHPRAH	61
Db	3	TTTTSHANTWTL-----	TKTKLMSPSKSEFEGPASNISILNHOHNROSLOINSSLOAPILH	57
Oy	62	FPKOSSN--SPA--IYVKPAKSEN-----	TKOMNLFORAAAALDAAGFLVSEHKLP	112
Db	58	FPKOSSNTQTPKNNTISHPRPOENNSSSST	SKWNLVOXAAAMALDAVSALTKELEHR	117
Oy	113	LPKTADDSVOLAGNFAPVNROPYARRULPVVGKL	LPDSIKGYVRANGANPLHEBVTGHFFED	172
Db	118	LPKRADRVOYSGNFAVPENPCVQSILPYTGK	IPKCVGVYVRNGANPLFEPTAGHHFFD	177
Oy	173	GDGWAHAHVKEBHGSASYACSFOTNRPEVOROLGR	PVPFKAIAGELHGHTGIARMLFYAR	232
Db	178	GDGWAHAVOEFENGCSASACRFETTERLYEROKA	LGREVPFKAIAGELHGHSIGIARMLFYAR	237
Oy	223	AAAGIYVPAHCTGYANNGLTYFNGRILLAMEBDLL	PVOYTTPRGDLKITYGRRDPFGOLES	292
Db	238	GLFLVLVDHRSKETGYAANGLVYFNKRLLAMESDDL	PYHVKYTPPGDLKTGTGREDFPGOLKS	297
Oy	293	TMIHAPKVDPESGEFLASTDYVASKPLTKFRFSP	GPQTSPDEIOLDOPTMHHPFATTE	352
Db	298	TMIHAPKPIDPVSGELFALSTDVIQKPLTKFRFS	KNGEKSNDVEIPEVEDPTMHHPFATE	357
Oy	353	NFVVVPPDOOVFXKLPENIRGSGEPVYVDKXNVAF	FGLIDYYAEBSNKIKWINDAPDCFCHL	412
Db	358	NFVVIPDDQVVFKEKSEMIRGSGEPVYVDKXNV	SFGLDKXYADGSLKKMVAEVPDFCFCHL	417
Oy	413	WMANEPEPTDEVVIGSCMPDPDISFNESDENLK	VYLSEIRLNLTKGESTRRPIINSEDQ	472
Db	418	WMANEPEPTDEIVIGSCMPDPDISFNECDEGLK	VYLSEIRLNLTKGSTRKSIITENDE	477
Oy	473	QVNIIEAGVNRNMJGRTKTRAYIALALAPMKV	SGFKADVLTITGEVKKHLJGNNRYGGEPL	532
Db	478	QVNIIEAGVNRNMJGRTTEYAYIALIEPMFKV	SGFKAVNLFTEGEVKFIYGNCKKGGEP	537

QY 533 FLGEG--GGEDDEGYILICEFVDEKTEKWSLEOIYVAWSLEWATYKPLSRPYGPHGFIIG 590
Db 538 FLPRDPNPKSEDDGIIILAFVDEKEMKSELOIYVAWSLKEATYKPLSRPYGPHGFIIN 597
QY 591 ADDILAKQ 597
Db 598 ANDILANQ 604

RESULT 3
E96812
protein F3F9.10 [imported] - Arabidopsis thaliana
C.Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

R:Rthologists; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,ansen, N.F.; Hughes, B.; Hulzar, L.
N:Native 408, 815-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marziani, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallio,ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A:Reference number: A86141; MUID:210156719
A:Accession: E96812
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-657 <SMD>
A:Cross-references: GB:AE005173; MID:98052533; PIDN:AAF71797.1; GSPDB:GN00141
C:Genetics:
A:Gene: F3F9.10
A:Map position: 1

Query Match	68.9%	Score	2170.5	DB 2	Length	657
Best Local Similarity	67.7%	Pred. No.	1.3e-156			
Matches 419: Conservative	80	Mismatches	87	Indels	33	Gaps
						9

```

0Y 1 MASFATATAAGRNLT-----GCHHTQPPRLSSOSSDLSYCSLPE-----MASRVT 45
Db 51 MASTLLLEPSTQJEDLDRTESSSSSRKLOS-----LPSSTLRNKLKLVPCYSSSVN 105
0Y 46 RKLNVSSALHT-----PRALHPKQOSSNSPAIVKPKAKESNTKOMLEFORAAALDAAE 101
Db 106 KKSVSSSLQSPFPKPSPM---KTLCDVNTLI-PRKTNQNP-KLNPQSTAAWLDAYE 160
0Y 102 GFLVSHF-KLHPLEKTDAPSOVQIAGNAPAPVNEQVYRRNLLPVYGLKPLDSIKGYVVRKANP 160
Db 161 NAMLSHERRRHPHRTADPAVOIAGNFPFEKPKVYVNLPLTGVPRCIGGVYVRKANP 220
0Y 161 LHEPVYTHHFEDGDMVAIVKFEHGSSASYACRFQTNREVOEROLGRPRVPAKALGELGH 220
Db 221 LHKPVSGHLLDGDGMVAIVAFDNGSVSYACRFETNRNLVQERCGRPVPAKALGELGH 260
0Y 221 TGIARLMLFYAPAAAGLYDPAHGTGVANAGLYVNRGLLMASEDDLPYQVQITPNEGDKT 260
Db 281 LGIAKLMLFENRRGLFGLVDPRGGVGNAGLYVNRGHLLMASEDDLPYHVKVQITQGLET 340
0Y 281 VGRPFEDQLSTMIAPHKVPYPEGSELFAISDYVVSAPYLKYPFSPDPGRKSPVEYOLD 340
Db 341 SGRIYFDQQLSTMIAPHKIDLPETRELFAISDYVVSAPYLKYPFETSDGKSPDEYELPD 400
0Y 341 OPTMMDHAITEENFVVDQOQVFEKPEMIRGSSPVYVDKNKVARFGILDKYAEDESSNIK 400
Db 401 OPTMIDHAITEENFVVIDQOQVFEKPEMIRGSSPVYDEKKSRFILLNKNNAKDAASSIQ 460
0Y 401 WIDADDCCFHLMANMEPPETDEVYVIGSCMTPPDSTIFNEDEDLKVLSLRNLKTGE 460
Db 461 WLEVPDCCFHLMNSWEPPETDEVYVIGSCMTPPDSTIFNEDETLQSVLSLRNLKTGE 520
0Y 461 STRRPIIENEDQVLEEGVNRNMLGRKTFATVLAIAEPKPKSGAKVDLTGTGEVKKH 520

```


Db 521 STRRRVVIS --- EQVNLLEGWVNRNLLGRKTRKAYLALTEPMKVSQFAPKYDSTGEIRKY 577

0Y 521 LYGNRRYGGEPFLFPGEGEEDGYIILCFVHDEKTKWSELOIVNAVSVLEAVTKLPSRV 580

Db 578 IYGEKRYGGEPLFLTPSGGDEEDGYIMVFVHDEKRVKSELQINAVNMKLEATVTLPSRV 637

0Y 581 PYGFGHTFICGADDLAKQVY 599

Db 638 PYGFGHTFISKEDLSKQAL 656

RESULT 4
AB6425
Probable 9-cis-epoxycarotenoid dioxygenase [imported] - Arabidopsis thaliana
C.Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
R.Accession: AB6425
R.Author(s): A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizar, L.
Nucleotide 408, 816-820, 2000
A.Author(s): Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A.Author(s): Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A.Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A.Reference number: AB6141; MUID:21016719

A:Cross-references: GB:AE005172; NID:g11094779; PIDN:AGC29711.1; GSPDB:GN001411
C:genetics:
A:Map position: 1

Query Match	64.9%	Score 204.5	DB 2	Length 589
Best Local Similarity	68.1%	Pred. No. 4.1e-147		
Matches 390	Conservative	69	Mismatches 101	Indels 13
				Gaps 4

```

QY 27 SOSSDSIYCCSLPMASRYTRKLNVSALHPALHPKÖSSNSPAIVKPKAKESNTKOM 86
      | | : : : | | : : | | : : | | : : |
Db 27 SPSSSVSTNTNPKR-----RRKLSANSVSDPNLNLNPNYPSPNPPII-----PEKDTSRW 76

```

QY 87 NLFQRATAAALDAAEGFLVSHKEKLPKPKADPSVQIAGNFAFVNEQPPRRNLPPVGGKLP 14
| | | | | | | : : | | | | | | | : : | | | |
Db 77 NPLQRRASAAIDFAETALRRERSKPLPKYDPRHQISGNYAPVPEQSVKSSLSVDGKIP 13

QY 147 DSIKGYVVRNGANPLHEPVTGHHFFDGDGMVHAVKFEHGSASACRFQTNRFRVGERQLG 20
 | | | | | | | | | | | | | | | | | | | | | | : | | | | | | | | | | | | | | | | | |
 Db 137 DCIDGYVLRNGANPLFEPVSGHHFFDGDGMVHAVKITNGDASYSCTFETETRLVGEKQLG 19

197 SPIFFAIGELHGHSGIARLMFVARGLFGLLNKNGTGVANAAGLVYFHDRLANSEDDL 25
 20/ RPVFPFAIGELHGHGTIGIARLMFVARAAAGIVDPAGTGTGVANAAGLVYFNGRLANSEDDL 26
 1:|||||:|||||:|:|:|||||:|||||

207 FIVQVGLIFNGDNLVGRKDFDQVLESIMVHPVADPSSGELFALSIDVSNPILKIFKFS 32
 |||:::||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||
 Db 257 PYQVRATDNGDEETIGREDFDGQLSSAMIAHPKIDPVTKEFLALSYDVKKPKPLYLKYKFS 31

```

317 PEGEKSPDVEIPLASPTMMHDFALTENFVVPDQVVFKLSDMFLGKSPVKYDGEKISR 37

```

```

377 GILPRAKADASEMWVESPETFCHILMNANESPETIDEVVISSCMTPADSIFNEDEQNL 43

```

```

|||||: || | :|||:| |:|:|
Db 437 SVLSEIRLNKTKGSKSTRITIPG-SVQNMLEAGMVNRLGKRTRYAYLAIAEPMPKVSG 49

```

```

Oy 507 FAKVOLLITGEYKKHIXGNRRGGEPFLPQ--EGGEPEGYIILCFYDEKTKWSELOIVN 564
Db 496 FAKVLLSTGEYKNNHTTGGCKTGGEPFLPRGLSESGEDDGTITMSFHDSESESLHIV 555
Oy 565 AVSLEVEATVKLPSRNPYGGFHGTITGADLAKQ 597
Db 556 AVTLELEATVKLPSRNPYGGFHGTITFVNSADLMQ 588

```

RESULT 5
T04531

nine-cis-epoxycarotenoid dioxygenase homolog F26J12.10 - Arabidopsis thaliana
N:Alternate names: hypothetical protein T9A21.200

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999

C;Accession: T04531; T04937
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft,

submitted to the Protein Sequence Database, February 1998
A: Reference number: 215377

A;Accession: T04531

A; molecule type: DNA
A; Residues: 1-583 <BEV>

A; Cross-references: EMBL:AL021710
A; Experimental source: cultivar Columbia; BAC clone F28J12

R; Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K. submitted to the Protein Sequence Database, February 1999

A:Reference number: Z15390
A:Accession: T04937

A: Molecule type: DNA
A: Residues: 1-377 <E

A;Cross-references: EMBL:AL021713

C;Genetics: A;Experimental source: cultural Columbia; bac clone 15A2

```
A;Map position: 4
A;Note: F28J12.10; T9A21.200
```

Query Match	63.2%;	Score 1991;	DB 2;	Length 583;
Best Local Similarity	64.5%;	Pred. No. 4.7e-143;		
Matches 380; Conservative	82;	Mismatches 105;	Indels 22;	Gaps 6

```

Oy 16 GGNHTOPPLSSQSSDLSY--CSLPMASRYTRKLNVSALHTTPPALHPEPKQSSNSPAIV 73
    || | || : | : : : | | : : |
Db 11 GGIKTWPI---QAQIDLGRPIRKQPKVIKCTQVIDYTE-LTKRKQLFTPTPTTAPP--- 62

```

```

Oy 74 VKPKAKSNTKQNLFORAAAAALDAEGEFLVSHEKHLPPLKTADPSVOIAGNEAPVNEQ 13
      : | ::+:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 63 -----QHNPRLNIFQKAAAIAlDAERALISHEQSPPLKPTADPVOIAGNYSVPES 11

```

QY 134 PVRRLNLEPVGKLPDSIKGYVVRNGANPLHEPYTGHHFFDGDGMVAHKFEHGSASTACRF 13
 117 SVRRNLVEGTIPDCIDGYIRNGANPMFEPTAGHHLFDGDMVAHKLTNGSASTACRF 17

```
QY      194 TQTNRFRQERQLGRPVFPRAIGELHSGTGLRLMLFARAAGLVDPAHGTVANAGLVY   25
        | : | | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db      177 TKTERLYQEKRLLGRPVFPKAI GELHSGIARLMFLFYARGLCGILNNQNGVAVNAGLVY   23
```

```

QY 254 FNGRLTMSSEDDLPYQVLTTPNGDKTVGRKDFDQGLSTIMIAHPKVDSESGELTALSVD 31
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 237 FNNRLTMSSEDDLPYQKLTITQTDGLQTVGRIDFDGQLKSAMIAHPKLDPVTKETALSYD 29

```

Db 297 VKKPKLYKFRSPDGVKSPELEIPLPTMIDFAITENFVIRPDQGVFKLGEMISGK 35

```
074 GFVILNNVNANGLDCKTHEDSSNLMIDFDCFEHMMNADEFEIENVJSCMIF 4  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :  
Db 357 SPVFDEKVSRLGIMPKDATEAGLIWNSPETCFEHLMAVESPEETEIVIVISSCMSP 41
```

```

417 ADSIFNERDELSRVLSEIRINLRPRKTTIRSSLVNEQ-VNLEIGMVRNRLGKRTFA 47

```


C.Species: Arabidopsis thaliana (mouse-ear cross)
C.Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C.Accession: T49193
R.Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Rudd, S.; Lemcke, K.
submitted to the Protein Sequence Database, April 2000
A.Reference number: Z25018
A.Accession: T49193
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-538 <R1E>
A.Cross-references: EMBL:AL163818; GSPDB:GN00061; ATSP:MAA21.150
C.Genetics:
A.Gene: ATSP:MAA21.150
A.Map position: 3
A.Introns: 72/3; 96/2; 101/2; 138/3; 172/1; 193/1; 230/1; 291/3; 338/1; 377/2; 411/2; 49

Query Match 29.8%; Score 938; DB 2; Length 538;
Best Local Similarity 37.0%; Pred. No. 3.8e-63;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;

QY 64 KSSNSPAIYVKKR-AKESNTKQNLFOFAAAALDAAGFLVSHKHLPLKPTADPSVQ 122
DB 4 KLSDGSIIIVHRRPSKGFSSKLIDLRLVYKLT-----HDASLPILH-----Y 47
QY 123 IAGNFAPV-NEOPVRRNLPVVKKLPDSIKGYVRKNGANPLHEPTVGHNFDDGGMVAHK 181
DB 48 LSGNFAPIRDETPPVKKLPVHGFLPECLNGEFGVRRGPRPKFDVAGYHMFDDGGMHGR 107
QY 182 FEHGSASYACRFOTNRFVQERQLGRVYFPAKIGELHCHTGIARLMLFARAAGLIVDPA 241
DB 108 IKDGKATVYSRYVKTSLKQEEFGAAKFMK-IGDLKGFEGILMVNVQOLRTKLKILDMT 166
QY 242 HGTGVNAGIYVYNGRLLAASEDDLPRVOYOTPNGLDKTVGGBFDDGOLESTMIAPKVD 301
DB 167 YGNGITNTALVYHNGKLLALQADKPRYIKVLEDDLTGLITIDYDKRLTHSTFAPKVD 226
QY 302 PESGELFALSYDVVSKPELYKFRFSPDGTSPDVEIQLDQPMHDFATLENVVPVDDQ 361
DB 227 PYTGEMETFEYS-HTPPYLYRVATSKDSIMHDPVITITSEPIMMDFATLETYATFMDLP 285
QY 362 VYFKLPEMTIGGSPV-YDKNKVAREGILDKYAEEDSSNIKWIDAPDCFCFHLMAEEPE 420
DB 286 MHRPRKEMVKEKKMIYSEFDPKKARFGVLPYAKDELMIWMELPNCFFHNAWME-E 343
QY 421 TDEYVYIGSMTPTD-----SIFNESDENLKSVLSEILNLKTGSESTRPIISNEQOVN 475
DB 344 EDEYVYLITCLRNPDLDLVMSGKVKLENGENLEYEMFRNMKTGSASQKLSA----- 398
QY 476 LEAGVNVRRNLGKRTKPAVALALAEPPKVSQFAVD-----LTTGEYKKNILY-- 522
DB 399 VDEPRINECTYTKKORVYVIGTILDSIAKVTGIILKFDLAAEATGKRMLEVGGIKRITYL 458
QY 523 GDNRYGGEPLFLPEGGEDEBGYLLCFVHDEKTYKSELQIYNAVSLEVE--ATVKLPSRY 560
DB 459 GEGRYGESEALYVPRTEEDDGYLIFEVHDENTGKSCVTVIDAKTMSAPRAVAELPRHY 518
QY 581 PYGFHGTFIGADDLAKOV 599
DB 519 PYGFHALFVTEEDLQEQTL 537

RESULT 9
T51734
neoxanthin cleavage enzyme [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51734
R.Neill, S.J.; Burnett, E.C.; Desikan, R.; Hancock, J.T.
J. Exp. Bot. 49, 1893-1894, 1998
A>Title: Cloning of a wilt-responsive cDNA from an Arabidopsis thaliana suspension culture
A:Reference number: Z24454

A:Accession: T51734
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A.Residues: 1-538 <ENB>
A.Cross-references: EMBL:A005813; PIDN:CNA06712.1
A.Experimental source: cultivar Landsberg erecta
C:Genetics:
A:Gene: ncl

Query Match 29.7% ; Score 937; DB 2; Length 538;
Best Local Similarity 37.0%; Pred No. 4.5e-63;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;

QY 64 KSSNSPAIVVPR-AKESNTKOMFLPRAAAAALDAEGEVLVSHEKHLLPLPTADPSVQ 122
| | :
Db 4 KLSDSIIISVHPRRSGFSSKLDDLRLVVKLM-----HDASLPILH-----y 47

QY 123 IAGNFAPV-NEOPVRNNLPVGKLRPDSTIKGVYVRNGANDLEPVTHGHHPFDGDNVAHV 181
:
Db 48 LSGNFARPIDEDFRPVKDLRFHGELPECLNGEEVAVRGVPNKFEPAVAGVHWFDGDGMIGHVR 107

QY 182 FEHSASVACRTQTQNRFRVQRQLGRVPERKAIGELHGTITARIMLFYARAACIOTPA 241
:
Db 108 IDGGATTVSRKYTSRKLOEFEPGAANKFMK-TGDLKGFGGLMNVNIQOLTKYLIDMT 166

QY 242 HCTGYANAGLVFNCRLLAMSEDDLPVOYTIPNDLKTVGRRFPDGOLESTMIAHKPYD 301
:
Db 167 YENGANTPALVHNCKILLALOGEADKPVIKVEDDDLOTLGILDOKRLTSHFTAHKPYD 226

QY 302 PESSGFALSTDVVSKPKYIKFRFSPODTGSVDVEIQLDQPMNHDAITENEVVFVDPQQ 361
:
Db 227 PYTGMEFFTFGVYS-HTPRYLTLYRVISKDGIMBDPVITITSEPIMMDFAITETAYAFMDLP 285

QY 362 VYEKLPEMRIGGSPVY-YDKNKVARFGILDKYAEDSNINMKWDAPDCCFHMMNWEDPE 420
:
Db 286 MHPREREWEKKRMKIYSDPTKARKPGVLPRAKDELMIWFELRCNPFHANWKE--E 343

QY 421 TDENVVIGSCMPDP-----SIFNESDENLKSVLSEIRLNLTGSTSTRPIISNQDOVN 475
:
Db 344 EEDEVLLTCRLRNPLDLLDWMSGKVKKELLENFGNELTEMFRNMKTSASQKILLSASA----- 398

QY 476 LEAGVANRMNLGRKTRERVALLAERPWPVSGFAKVD-----LTGGEVKNHLV-- 522
:
Db 399 VDFPINECYTGTKORYYYVTGLIDSIAKVTGIKFIDLHAEAETGRKMLEVGINIGIYDL 458

QY 523 GNRRYGGERLFLPGSGGEDBGYIICFVHNDEXTWKSELQIVANAFLVE--ATVKLPSRV 580
:
Db 459 GEGRGSESLVYPRETAEDDGYLIFFVHNENTGKSCVTVIDAKTMASPAVAVVELPHRV 518

QY 581 PGFHGTFPGADDLAKOVV 599
:
Db 519 PYGFHALFYTEEQLOEQTL 537

RESULT 10
AG1944
hypothetical protein al11106 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A>Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AG1944
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigun-
Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabatake,
DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1944
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-475 <KUP>
A:Cross-references: GB:BA000019; PIDN:BAB73063.1; PID:g17130452; GSPDB:GN00179
A:Experimental source: strain PCC 7120

C:Genetics:
A:Gene: all1106

Query Match 23.7%: Score 745.5; DB 2: Length 475;
Best Local Similarity 34.5%: Pred. No. 1.3e-48;
Matches 169; Conservative 91; Mismatches 171; Indels 59; Gaps 15;

```
QY 123 IAGNFAVNEQFVRNLPVVGKLPDSIKGVYRNGANPLHEPTGHHFPGDGMVAVKF 182
DB 22 LEGNFAVNEHEITTDLTKYIGELPRLSGMFVRNGNPNPMTPIGQVHNMFDGDMHGVRI 81
QY 183 EHGSAASYACRTQTNRFVEORQLGRPEPKAIGELHGTHTGAKRLMLFYRAAGIVD--- 239
DB 82 SNGKATYRRNRYVTOFMQJHEHGAQIW-----TGLEPPQ 117
QY 240 ---PAHGTVANAGLVYFNGRLLAMSEDDLRYOVQITPNGDLKTVGRPDQLESTMA 296
DB 118 TELPSRNTG--NTALIMHAGQLLALMEGAPYAIQVP---DLASIGETYNNQDLSAFTA 172
QY 297 HKRVPESEGLFALSYDVYSKPYLKFRSPDGTSPDVEIOLDOPTMHDFAITENFVY 356
DB 173 HKRVDPVTGEMMFEGS-FAPRYLHVSYSATGELVRYPIIDLPKGVMMHDAITANTYTI 231
QY 357 VPDQOVFKLPEMIRGSPVYVDKKNVAFGLDKYAEDSSNIKWIDARDCFCFLMNM 416
DB 232 FMDLPLTFSEVRMQRCEPMLMEFSDRSPRFGILPRHG-DNSQIRMFEDRSCYVHTLMAY 290
QY 417 EEPPEDEVYVIG-----SCMTPPDSIFNESDENLK-SYLSERLNLKGTGSTRRIISN 469
DB 291 E--DDDEVLLFACPMKSTVILASPS--QTPDEADIPRLHRMRFLLKGVKHEEML--- 342
QY 470 EDQOVNLEAGVNRNMLGKRTKFAVLA-LAE-PWPKVSGFAKVDLTGVEKKHLYDNRY 527
DB 343 --DDVASEPRRINENLTGPTQYGYTSRLAKGSIRPLEBILIKYDLSNASKSÖNTEYOGK 400
QY 528 GGEPLFLPEEGE-EDEGYILCFVHDEKTKWSELOIVNAVSELE--ATVKLPSSRVYGE 584
DB 401 GSEAVFVPRPAGATVEDDGWILTYVYDTGSESELVYINADINSEPIARILLPQRPVYGF 460
QY 585 HGTFIGADLL 594
DB 461 HGIVWTEQL 470
```

RESULT 11
T17019

hypothetical protein - apple tree
C:Species: Malus domestica (apple tree)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T17019
R:Matillon, B.; Kettmann, R.; Arredouani, A.; Hesquet, J.; Boxus, P.; Burny, A.
Plant Mol. Biol. 36, 909-915, 1998
A:Title: Apple messenger RNAs related to bacterial 11gnostrilbene dioxygenase and plant S
A:Reference number: Z18655; MUID:98179104
A:Accession: T17019
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-446 <MAT>
A:Cross-references: EMBL:Z93765; NID:q2924324; PIDN:CA00784.1; PID:q2924325
A:Experimental source: cv. McIntosh, strain Wj1c1k

Query Match 23.7%: Score 745; DB 2: Length 446;
Best Local Similarity 39.8%: Pred. No. 1.3e-48;
Matches 150; Conservative 76; Mismatches 129; Indels 22; Gaps 9;

```
QY 231 ARAAGIYDPAHGTVANAGLVYFNGRLLAMSEDDLRYOVQITPNGDLKTVGRPDQLE 290
DB 76 ARVLGQVYRPAANGIGLANSLAFGDRYLALGESDLPYSLRLTNSGDDITFLGHDPDKGL 135
QY 291 ESTMAHKKVDESGELFALSYDVYSKPYLKFRSPDGTSPDVEI-OLDOPTMHNDRA 349
```

```
DB 136 SMMMTAHPKIDBDTGEAFARFYGEI-RPFLTYFRDPSNGVKOPDVFISMTPTFLHDA 194
QY 350 ITENFVYVDDQOVVFKLPFEMI-RGSPVYVDKKNVAFGLDKYABDSSNIKWIDARDCFC 408
DB 195 ITKKAHIFADIDQGLNLDIMTKRATPFGDPSKVRIGIVIPLYAKADESMKFEVPGN 254
QY 409 CFHLMNAMESPETDEVYVIGSCMTPPDSIFNESDENLKVSLEIRLNLKTGSTRRPIIS 468
DB 255 GVHATNAMD--EDDALVMVAPVNLASAHLERYD-LVHCLTEKVRIDLKTGTYTRQPIST 311
QY 469 NEDQOVNLEAGVNRNMLGKRTKFAVLA-LAEPPKVSFGAKVDLTGGE-----YKHLYG 523
DB 312 R-----NDFEAVINPAVYLRKKNVYVAEGDPMPKISGVVKKLDVSNVEHKECTIVASRMG 366
QY 524 DNRVGEPLFLGE-----GGEDEGYILCFVHDEKTKWSELOIVNAVVS--LEVEATVKRP 577
DB 367 PGCYGGEFFVAREPNPNADEDNGFLVSYVHDEKAGESRFLYMDAKSQDLIVAAYRMP 426
QY 578 SRVYGFHGTIGADLL 594
DB 427 RRVYGFHGLFRESDL 443
```

RESULT 12

E87345
conserved hypothetical protein CC0776 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87345
R:Nietman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87345
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-483 <STO>
A:Cross-references: GB:AE005673; NID:q13422015; PIDN:AAK22761.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0776

Query Match 16.8%: Score 528.5; DB 2: Length 483;
Best Local Similarity 30.9%: Pred. No. 4e-37;
Matches 160; Conservative 89; Mismatches 203; Indels 65; Gaps 22;

```
QY 108 EKLHPLPKTADPS--VOIAGNFAVNEQFVRNLPVY-GKLPDSIKGVYRNGANPLHEP 164
DB 4 ERLPVRTSLGPTNHPYMTGPMTQHEEVANAMDLEVLGATPADLDGVLVLRNTENPVHDP 63
QY 165 VGHNFHFDGDMVHAVKFEHGSASYACRTQTNRFVQEOQLGRPVPAKAIIGELHGTGTGA 224
DB 64 IGRYHFDDGDMIHQIEFGGAATYRNRPRFCEAEDEVNEGM---GGIMDGPVGS 119
QY 225 RLMLFYARAAGIYDPAHGT--GVANAGLVYFNGRLLAMSEDDLRYO---VOITPNGDL 278
DB 120 K-----RPGFG---AHALNDSASTDIVNNGEALIAF-----YQCEAATRLDPL-TL 163
QY 279 KTVGRFEDGQLESTWIAHPKVDPESGELFALSYDVYSK-PYLYKFRSPDGTSPDVE 336
DB 164 ENIGVASW--APLEG-VSAPRKVDATGELMFNY---SKAMPYMHGVYVDPGKKRVYQG 218
QY 337 IOLDOPTMHNDPALTENFVYVDDQOVVFKLPEMIRGSPVYVDKKNVAFGLDKYAEDS 396
DB 219 VPLPGGRPLPHDAFSSKYAILNDLPVFMQDELMARIDHVAHRLHKGIPSRFALVPR---EG 275
QY 397 SNIKWIDARDCFCFLHNMAMESPETDEVYVIGSCMTPP-----DSIFNESDEN 444
DB 276 GERPMFEAPRYVVLHNLNAYE---DGDEVYLDGYFOEKRPRLRELEGARPDGHMLAYLDH 333
QY 445 LKSVLSEI--RLNLKTGSTRRPIISNEDQOVNLEAGVNRNMLGKRTKFAVYLAAP- 500
```

```

Db      334  --SELPKLLHRMFNFNLKGTETTERKL-----DDRV-LEFGFNFKYAKGKPRRYVSTTARG 386
Oy      501  WPKVSGFAKVDLTITGCEKHLHYGDNRYGGEPLFLPDEGG-EEDEGIIICEVHDEKTMXK 559
Db      387  WFLFNGFVKHDIETGESSIALPEGRYASEAFPAKVGAVDEDDGYLVSFFIDENKGASE 446
Oy      560  LQIVANVSLVEATVTKLPSRVYPGFHGFIGADDLAK 596
Db      447  CLIVDAKRFEEVYCRIALPKHLSSGTHYWAGREMLTK 483

RESULT 13
A70534
hypothetical protein RV0654 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70534
R:COLE, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: A70534
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-501 <COL>
A:Cross-references: GB:295972; GB:AL123456; MID:93261790; PIDN:CAB09380.1; PID:e319166;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV0654

```

DB 447 AAEAEDGILMGYGMHGRDEGOLLILDAQTLESIAIVHLPQKRVPMFGNGW 497

RESULT 14

A70582

hypothetical protein RV0913c - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: A70582

R:COLE, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, R.; Jandaurem, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

A:Reference number: A70500; MUID:96295987

A:Accession: A70582

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-502 <COL>

A:Cross-references: GB:265210; GB:AL123456; MID:g3261757; PIDN:CAB08511.1; PID:c3152121

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV0913c

Query Match	14.48;	Score 454.5;	DB 2;	Length 502;
Best Local Similarity	27.68;	Pred. No. 1.7e-26;		
Matches 133; Conservative	73;	Mismatches 190;	Indels 101;	Gaps 18;

QY 141 VGLRLPDSIKGVYVRNGANPLHEEVTGHHFEDGDMVHAKVFEEGASASYACRETQTNREV 200
1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : :
Db 41 VTGEVPADDDGIYIRNTENPLHPAEATYRPEDGSGMIHVGGFRGKAFAFRNFRITDGL 100
1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : :

```

QY      201 QEQQLGRVPFPAKIGELHCHTGLARLMLEYAR----AAAGIYDPAHGTGV-ANAGIV--- 252
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      101 AENEAGGRPLP-----GLAEVQLAKREHGWGARGLKKDKDASIDIVIRGIALTS 150

```

[illegible]

```

Db      201  YS-KQPPVRYRGVNDONNELHYHVDPLGPRFLPHDMATTEYVILND----- 247
QY      372  GGSPPVYDKKNAAREILDKY-----AEDSSNKIKWDAPPCPFHLMNAEPE 420

```

```

Db      248  --FPLEFMDRLLEEDVHDLRFREYLPSPFAVVAARRKNDIRMEADPTEFLHPTNAE--Q 303
Qy      421  TDEVYVSGCMTPPDGIFNESD-----ENDKSVLSEIRLNKJTE 460
          :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      304  GDEIVVDG-----FEEGDQPLPDGCTKWEKLFRELRALDRQSLRRRRRLNNVTC- 353

```

Qy	461	STRRPILNSDDQVN--LEAGVNVNNMGRKTKFAYLLAEP-WPKYSGFAKVDLTGGE	516
		: : : : : : : : : : : : : : : : : : : : : : : :	
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	354	-----AVHEEQLSSTIEFGTINADYAASSRYRYATATGPRSWFLFDGLVKHDLDTGN	406
		: : : : : : : : : : : : : : : : : : : : : : : :	
Qy	517	IKKNTKSNHDEQSLSEETFGESGE--EDKVVLTGKNDVETGSEETQTVNNGSTNIPKATG	575

QY	576	--LPSRVPCGFHCTFGADDLAK	596
DB	407	HECYSGDGDVYGSERMAARVSSASADDGLVLT	466
QY	576	--LPSRVPCGFHCTFGADDLAK	596

DB 467 LALPERISSGTHSAMVPGALRR 489

RESULT 15

S76169

hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2002, 10:12:13 ; Search time 13.49 Seconds

(without alignments)
1719.274 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTATAVSGRWLGNGNHT.....VPYGHGCTIGADDLAKOV 599

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	3.6	870	1 YCS0_YEAST	P25623 saccharomyc
2	113	3.6	466	1 DNAA_PROM1	P22837 proteus mix
3	107.5	3.4	658	1 PAK1_SCHPO	P50527 schizosacch
4	106	3.4	3562	1 PGCY_CHICK	O90953 gallus gall
5	105	3.3	783	1 CAD5_MOUSE	P55784 mus musculi
6	105	3.3	2504	1 FAS_HUMAN	P49927 homo sapien
7	104.5	3.3	1079	1 APCE_FREDI	P16566 tremyella d
8	103	3.3	775	1 SYPB_AQUAE	O67620 aquifex aeo
9	102	3.2	2493	1 CYAA_USTMA	P49606 ustilago ma
10	101.5	3.2	520	1 GLGS_BRANA	O9m462 brassica na
11	101.5	3.2	565	1 SCRL_SCHPO	O14335 schizosacch
12	101	3.2	512	1 GLGT_VICFA	P52417 vicia faba
13	100	3.2	3770	1 ACVS_EMENT	P27742 emeticella
14	99.5	3.2	383	1 PHYC_BACSU	O31097 bacillus su
15	99.5	3.2	1256	1 ATL_STRAU	P52081 staphylococ
16	98	3.1	741	1 MASZ_MYCTU	O50596 mycobacteri
17	98	3.1	1788	1 YP72_CAEEL	O09921 caenorhabd
18	97.5	3.1	518	1 CP74_ARATH	O96442 arabidopsi
19	97.5	3.1	691	1 SLP1_YEAST	P20795 saccharomyc
20	97	3.1	636	1 KDPG_SCHPO	O14019 schizosacch
21	96.5	3.1	426	1 UCR2_SCHPO	P39581 bacillus su
22	96.5	3.1	503	1 DLFA_BACSU	P32567 saccharomyc
23	96.5	3.1	862	1 SMP2_YEAST	O30408 b tyrocidin
24	96.5	3.1	3587	1 TYCB_BACBR	O92265 penicillium
25	95.5	3.0	865	1 NRPA_PENUR	O92265 penicillium
26	95.5	3.0	918	1 DNL1_RAT	O92265 penicillium
27	95.5	3.0	1144	1 DP3A_NEIMB	O92265 penicillium
28	94.5	3.0	590	1 HMAA_DROME	O92265 penicillium
29	94.5	3.0	638	1 GHR_RAT	P16310 rattus norv
30	94.5	3.0	953	1 LRAB_PASHA	P55118 pasteurella
31	94.5	3.0	1046	1 PPOC_WEITHE	P96177 weissella h
32	94	3.0	532	1 P72_MYCMY	P55801 mycoplasma
33	94	3.0	919	1 Y893_HUMAN	O94967 homo sapien

34	93.5	3.0	1275	1 GNRD_HUMAN	O13972 homo sapien
35	93	3.0	3067	1 CALC_MOUSE	O60847 mus musculi
36	92.5	2.9	354	1 OTX1_HUMAN	P32242 homo sapien
37	92.5	2.9	955	1 LRAB_PASHA	P55117 pasteurella
38	92.5	2.9	1057	1 SPST_CITON	O22060 citus unsh
39	92.5	2.9	1116	1 SLPH_BACBR	P38538 bacillus br
40	92.5	2.9	1131	1 CHSL_YEAST	P08004 saccharomyc
41	92.5	2.9	4829	1 BIR6_HUMAN	O9m09 homo sapien
42	92	2.9	880	1 CADE_XENLA	P33148 xenopus lae
43	91.5	2.9	474	1 GSA1_ARATH	P42799 arabidopsi
44	91.5	2.9	744	1 HXCI_HAEIN	P44523 haemophilus
45	91.5	2.9	896	1 APCE_SYNY4	O02907 synechocyst

ALIGNMENTS

RESULT	ID	YCS0_YEAST	STANDARD:	PRT:	870 AA.
AC	P25623	YCS0_YEAST	P25622:		
DT	01-MAY-1992	(Rel. 22, Created)			
DT	15-DEC-1998	(Rel. 37, Last sequence update)			
DE	16-OCT-2001	(Rel. 40, Last annotation update)			
GN	YCR030C	OR YCR30C/YCR29C.			
OS	Saccharomyces cerevisiae	(Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OX	Saccharomycetales; Saccharomycetaceae; Saccharomycaceae;				
NCBI_TaxID=4932;					
RP	SEQUENCE FROM N.A.				
RA	Cederberg H., Hohmann S., Schaaff-gerstenschlaeger I., Huse K.,				
RL	Zimmermann F.K.;				
RN	Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.				
RP	REVISIONS.				
RA	Gromadka R.;				
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.				
CC	- SIMILARITY: SOME, TO S.POMBE SPBC4C3.06.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: X59720: CNA42322.1: -				
DR	SGD: S0000626: YCR030C.				
KW	Hypothetical protein.				
SEQUENCE	870 AA: 96125 MW; 3F2CE2P1562E5277 CRC64;				
Query Match	3.6%; Score 114; DB 1; Length 870;				
Best Local Similarity	20.6%; Pred. No. 1.2; 195; Indels 180; Gaps 37;				
Matches 121; Conservative 92; Mismatches 180; Gaps 37;					
QY	3 SFPTATAVSGRWLGNGNHTPLSSOSSDLS---YCSLPMASRVTKLWVSAL--HTP 57				
DB	384 SSTATSIYK-----QRRRTYSSSKSNMTPPEASDPPLPLPHATPK-NVDAPVADTP 437				
QY	58 PALHF-----PKSSNSPAIVVKKAKESNTKOMNLFORAAAALDAAGFLVSHE 108				
DB	438 PAOTFPSEVPSPSTPOSS-----PTAKEDPSNL-----PKTVPIISISQIPLOPOS 485				
QY	109 KLHPLP-KTADPEV-----QIAGNFAPVMEOPRRRLPVVGKLPDSIKGVYRNGA 158				
DB	486 KTKPLPPEPSPISLPTATVDNOPSQ---VSRPLPHRAPA---LPSRKONFIHNRD 539				
QY	159 NPLHEPVYTGHHFDDGDMVAHVAFEHGASAYACRFOTNRFVDEROLGRVY---FPAKIG 215				

DB 540 SOLYSLP-----NHGSGA-----TPTSSLSIPDERPVSTLSSQITG 578
 QY 216 ELHGHTGIARLMEFYARAAGIVDP-AHGTVANAGLVYFNGLRLAMSEDDLPRYOVIPTP 274
 DB 579 ELKE-----LNPOATGSSSTISVSGSLFQHSGLDTSQGLNASTIAEVL 620
 QY 275 NGDLKTVGRPFDDGOLESTMIAPHKVDPESGELFALSY---DVVSKPYLKYFRFSPDGTK 331
 DB 621 NASEFK-----DGLMNSQOLI-----GEI-ALNYLPLNSVMSNPL-----PIG-- 655
 QY 332 SPVEVQLDQPTMHMDPAITENFV--VVPDQVYFKLPKEMIRGSPVYVDKKNKVARCITL 389
 DB 656 ---INRINNGAKFEKYLINQAFIERVAPBE---FKV-----NPSFIDSRTLGAI--- 699
 QY 390 DKYAEDESSNIKWIDAPDCFCFH-LMNAMPEPETDEVVIGSCMTPPSI----- 437
 DB 700 -KI-----SIKEPIAP--IYIHPVWR-FESHQASVVLTVKMSPSLPREISOYIEDLVVF 750
 QY 438 FNESEDNLKSVLSEIRLINLKTGST----RPII---SNEDQOV---NLEAGKVNRMNL 486
 DB 751 VNIDGAMATGALSQPGSFKSEKKRITWRFEKPVVLTFRNEGQRLIARFTIDGLAHBSAK 810
 QY 487 GRTKTFAYLALAE---PMPKVSQFAVVDLTGTGVKKHLXGDNRYGGE 530
 DB 811 GVYTKF---TISETDNALPHSGSGSGLTTCQELDE---NNPFGGE 851

RESULT 2

DNA_PROMI STANDARD; PRT; 466 AA.

AC P22837;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Chromosomal replication initiator protein dnaA.
 GN DNA.
 OS Proteus mirabilis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Proteus.
 NCBI_TaxID=584;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LM1509;
 RX MEDLINE=91033012; PubMed=2172087;
 RA Skovgaard O.;
 RT "Nucleotide sequence of a Proteus mirabilis DNA fragment homologous
 RT to the 60K-rnpA-rpmH-dnaA-dnaN-recF-gyrB region of Escherichia
 RT coli.";
 RL Gene 93:27-34(1990).
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION
 CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;
 CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS
 CC (DNA BOX): 5'-TTATTC(C/A)A(C/A)A-3'. DNAA BINDS TO ATP AND TO
 CC ACIDIC PHOSPHOLIPIDS.
 CC -1- SIMILARITY: BELONGS TO THE DNAA FAMILY.
 CC
 CC This SMS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M58352; AAA83958.1; -.
 DR PIR: JQ0733; IOEBV.
 DR InterPro: IPR001957; Bac_DnaA.
 DR Pfam: PF00308; bac_dnaa; 1.
 DR PRINTS: PR00051; DNAA.
 DR PROSITE: PS01008; DNAA; 1.
 KW DNA replication; DNA-binding; ATP-binding.
 FT NP_BIND 171 178
 SQ SEQUENCE 466 AA; 52974 MW; F8B67C142FE9FA41 CRC64;

Query Match 3.6%; Score 113; DB 1; Length 466;
 Best Local Similarity 21.9%; Pred. No. 0.56;
 Matches 84; Conservative 48; Mismatches 112; Indels 140; Gaps 21;

QY 105 VSEHKLHPLEKT-ADPSQIAGNEFAPVNEQPVRRNLPPVGLKLPDS-IKGYVRNGANPLH 162
 DB 83 VSAATTESVPEKTYTHPAV-----NSTPTNSQPVK---PSMDNQPSQLPELNTSRNSVNPKH 135
 QY 163 EPTVGHNFEPDGDGVNAVKFEHGSASACFTQTNRFQEROLG--RPVYPKAIGELHGH 220
 DB 136 K-----FD-----NVEGKSQOLAR-AAARQVANPGGAVNPLF-----LYGG 172
 QY 221 TGIARLMLFYARAAGIVDPAHGVV---ANAGLVYFNGLRLAMSEDDLPRYOVIPTNG 276
 DB 173 TGLGKTHLLH-----AVGNSIYERRKANAKVYVMH-----SRFQYDMVKALQNN 216
 QY 277 DLKTVGRPFDDGOLESTMIAPHKVDPESGELFALSYDVSKPYLKYFRFSPDGTKSPDVE 336
 DB 217 AIE-----DEK-----RYYR-SVDALLIDDIQ 237
 QY 337 IQLDQPTMHMDPAITENFVVPDQVVF---KLPKEMIRGSPVYVDKKNKVARCITDKYA 393
 DB 238 FFANKERSQEEFHTFVALLEGNOQILITSDRPKEING---VEDRLK-SRFG----- 286
 QY 394 EDESSNIKWIDAPDCFCFHLMNAMPEPETDEVVIGSCMTPPSIFNESEDNLKSVLSEI- 452
 DB 287 -----WGLVVAIEPPELETFRVAL-----LMKRADENOIOLPDEVA 321
 QY 453 -----RLNLKTGSTFRPIISN 469
 DB 322 FFIAKRLRSNVRELEGALNRVIAN 345

RESULT 3

PAK1_SCHPO STANDARD; PRT; 658 AA.

AC P50527;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Serine/threonine-protein kinase Pak1/shk1 (EC 2.7.1.-).
 GN PAK1 OR SHK1 OR ORB2 OR SPBC1604.14C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96112805; PubMed=8846783;
 RA Ohtllie S., Miller P.J., Johnson D.I., Creasy C.L., Sells M.A.,
 RA Bagrodia S., Forsburg S.L., Chernoff J.;
 RT "Fission yeast Pak1 encodes a protein kinase that interacts with
 RT Cdc42p and is involved in the control of cell polarity and mating.";
 RL EMBO J. 14:5908-5919(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Marcus S.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Beck A., Reinhardt R., Lyne M., Rajandream M.A., Barrell B.G.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 119-658 FROM N.A.
 RX MEDLINE=95320235; PubMed=7597098;
 RA Marcus S., Polymerio A., Chang E., Robbins D., Cobb M.H.,
 RA Wigler M.;
 RT "Shk1, a homolog of the Saccharomyces cerevisiae Ste20 and mammalian
 RT p55PAK protein kinases, is a component of a Ras/Cdc42 signalling

[illegible]

SQ SEQUENCE 3562 AA: 388078 MW: 98C56688C1602D2 CRC64;
 Query Match 3.4%; Score 106; DB 1; Length 3562;
 Best Local Similarity 22.7%; Pred. No. 36;
 Matches 118; Conservative 61; Mismatches 191; Indels 150; Gaps 29;
 QY 1 MASFAATA-ANSGRLGCGNHTQPLPSSQSSDLSLTCSSLLPMASTRKLNLSALHT--- 56
 DB 1231 MLSSESTGSSISLLTLGASPSQTEPGSGSISELEEVKIVKPSRATDXTVYISDLTSSSI 1290
 OY 57 -----PALHFPKQSSNSPAIVKPKAKSSNTKQML-----FQRAAALDAAEGL 104
 DB 1291 SAVDKIQGTSAKSPVSSKSPRIIT--PEDEVTSSDITIVIDESISPSKASAEDDLTKM 1348
 OY 105 VSHKILHLPK-----TADPSYQIAGN-FAPVNEQPVRRNRLPVGKLPDS- 148
 DB 1349 VEPE-----IDKEFYTSQATAVARPTAPRYMEATEALQPEVSPYTSH-----PDG 1396
 OY 149 --IKGVYRNGANGPLHEVYTG-HFFDGDGAVHAKFEGHSASVACRTQTNRVQERQ- 204
 DB 1397 TDIRLYVYQIQTGNPDTHDVNNEFLDFSRHILPHAVDETHTAESKQTEPCTSDSYDSS 1456
 OY 205 --LGRVPVPAKIGELHGTIGTARLMLFYARAAGIVDAHQGTGAN-AGLYVFNQRLAM 261
 DB 1457 YIILDPFP-----NFMDFEEEEE---DCENTTDVTTPPALPFINK----- 1495
 OY 262 SEDLPYVOQITPNGDLKTGVRGFPDQGLESTMIAPHVPEDESGEFLASDVVSKPYLK 321
 DB 1496 -----QQVTSAPKSTKAEARSD--QIES-VAHSK-NVTFQSQINTNFIISET--- 1540
 OY 322 YFRSPDGTKEPDV-----EIQLDQPLMMHDPALTENFVVVPPDQGVFKLPEMIRGSGP 375
 DB 1541 ---EASGTMPQSRKAGEWGAFEVYQPPA--DVAMLE-----P 1572
 OY 376 VVYQNKAKRARGILDKYVAEDSS-----NIKMIADADCQFHLMMAMEE---PETDEV 424
 DB 1573 VYSGESET---TTDKYLEITISVYEQSKKKEVY-----MMHGTESSTKDKTKNL 1620
 QY 425 VVIGSCMPPDSIFNSEDENLKSYSLEIRNLKTGESTR 464
 DB 1621 LLITNESSGDS--TESDLS-RSVPTET-LTMSHDESEK 1656
 RESULT 5
 CAD5_MOUSE STANDARD; PRT; 783 AA.
 ID CAD5_MOUSE
 AC P55284: 035542;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Vascular endothelial-cadherin precursor (VE-cadherin) (Cadherin-5).
 GN CDH5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 OX [1]
 RN RN
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain capillary;
 RC MEDLINE=96141083; PubMed=8555485;
 RX Bietler G., Breviario F., Cavada L., Berthier R., Schnuerch H.,
 RA Gotsch U., Vestveber D., Risau W., Dejana E.;
 RT "Molecular cloning and expression of murine vascular endothelial-
 cadherin in early stage development of cardiovascular system.";
 RL Blood 87:630-641(1996).
 [2]
 RP SEQUENCE FROM N.A. AND FUNCTION.
 RC STRAIN=BALB/C; TISSUE=Breast carcinoma;
 RX MEDLINE=97364256; PubMed=9220534;
 RA Matsuyoshi N., Toda K.-I., Horiguchi Y., Tanaka T., Nakagawa S.,
 RA Takeichi M., Immura S.;
 RT "In vivo evidence of the critical role of cadherin-5 in murine

CC	long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC	-1 CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA + acetyl-[acyl-carrier protein].
CC	-1 CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA + malonyl-[acyl-carrier protein].
CC	-1 CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) + [acyl-carrier protein].
CC	-1 CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] + NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC	-1 CATALYTIC ACTIVITY: (3R)-3-hydroxyphenylmethyl-[acyl-carrier protein] = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC	-1 CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC	-1 CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-carrier protein] + oleate.
CC	-1 SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC	-1 TISSUE SPECIFICITY: PROMINENT EXPRESSION IN BRAIN, LUNG, AND LIVER.
CC	-1 MISCELLANEOUS: THE RELATIVELY LOW BETA-KETOACYL SYNTHASE ACTIVITY MAY BE ATTRIBUTABLE TO THE LOW 4'-PHOSPHOPANTHETHEINE CONTENT OF THE PROTEIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: U26644; AAC50259.1; -.
DR	MIM: 600212; -.
DR	InterPro: IPR001227; Acyltransf_domain.
DR	InterPro: IPR002085; Adh_zn_family.
DR	InterPro: IPR000794; Ketoacyl-synt.
DR	InterPro: IPR003880; Phosphopant_attach.
DR	InterPro: IPR000051; SAM_bind.
DR	InterPro: IPR001031; Thioesterase.
DR	Pfam: PF00698; Acyl_transf. 1.
DR	Pfam: PF00107; adh_zinc. 1.
DR	Pfam: PF00109; ketoacyl-synt. 1.
DR	Pfam: PF02801; ketoacyl-synt_c. 1.
DR	Pfam: PF00550; pp-binding. 1.
DR	Pfam: PF00975; Thioesterase. 1.
DR	PROSITE: PS00606; B_KETOACYL_SYNTHASE. 1.
DR	PROSITE: PS00075; ACP_DOMAIN. 1.
KW	Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine; Hydroxylase; Oxidoreductase; Transferase; lyase; NADP; Pyridoxal phosphate.
FT	DOMAIN 1 413
FT	DOMAIN 428 815
FT	DOMAIN 1650 1857
FT	DOMAIN 1858 2113
FT	DOMAIN 2118 2174
FT	DOMAIN 2202 2504
FT	ACT_SITE 161 161
FT	ACT_SITE 580 580
FT	ACT_SITE 876 876
FT	NP_BIND 1666 1683
FT	BINDING 1699 1699
FT	NP_BIND 1880 1895
FT	BINDING 2151 2151
FT	ACT_SITE 2302 2302
FT	ACT_SITE 2475 2475
SQ	SEQUENCE 2504 AA; 273100 MW; 8AA9E809B2338DFA CRC64;

Query Match	3.3%	Score 105;	DB 1;	Length 2504;
Best Local Similarity	23.3%;	Pred. No. 25;		
Matches	81;	Conservative	41;	Mismatches 143;
			Indels	82;
			Gaps	19;

```

OY 11 SGRNLGGNHQTPPLSSOSSDLSVCSSTLPMASRVTARKLN--VSSALHTPRALHPKXS--66
Db 707 SARWL-----SITPEQOMHSSLRATSSEAYNNANLVSPVLQEOBALMHVEPAHV 756
OY 67 -----SNSPVAIVKKRAESNTWKQNMLQORAAAAAALDAEGFLYSHEKLHPDPPTADPSV 121
Db 757 LEIAPTPOQAVLKTKVRPDSCT----ILPRMKKHDRONLEFFLAGICGLHSJGIDAMPNA 812
OY 122 QIACGFAPVYNQOPVARNLPVYGKLPDSIKGYVVRNGANPLPEPTYGHFFPDGDGMHAVK 181
Db 813 I---FPPV-ESPAPRGTPLLSP-----IKMWSLAMDAAPA-AED 847
OY 182 FEHGSGAS-----YAC--RFOTQNRFCVEROL-GRPVFPKAIGELGHGTGIARLMLFARA 233
Db 848 PFMGSQSFSATITYCTPSSESPPDRVLVDHTJDGRVLPFAT-----GYLSIWYKTL--ARA 900
OY 234 AAGT-----VPAHGCVANAAGLVFNGRLLAMSEDLDPYOVQOTTPNGDKLTWGR- 283
Db 901 WAGLEQLPVVEDVYQHQAOTILLPKGTIVSLERVRL-----EATGAFEVSENGNLVSGKV 955
OY 284 RFDFOQLESTIMAHXK-VDPES--GEFLALSVDYVSKRYLKYFPFRSP 327
Db 956 YQMDDP-DPRLEDHDESPHNSPNSPIFLAQAEVYKEULRLGYDGP 1001

RESULT 7
ACE_FREDI ID ACE_FREDI STANDARD PR: 1079 AA.
AC AC PI6566;
DT DT 01-AUG-1990 (Rel. 15, Created)
DT DT 01-AUG-1990 (Rel. 15, Last sequence update)
Dr Dr 01-DEC-1992 (Rel. 24, Last annotation update)
DE DE Phycobiosome 120 kDa linker polypeptide, core (L-CM 92) (Core-membrane linker protein).
CN CN APCE.
OS Fremyella diploiphon (Calothrix PCC 7601).
OX Bacteria; Cyanobacteria; Nostocales; Rivulariaceae; Fremyella.
CX NCBI_taxid=1197;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RX RX MEDLINE=90192765; PubMed=2107546;
RA RA Houtmand J., Capriano V., Colombano M.V., Courain T., de Marsac N.: "Molecular characterization of the terminal energy acceptor of cyanobacterial phycobilisomes."; Proc. Natl. Acad. Sci. U.S.A. 87; 2152-2156(1990).
RT RT FUNCTION: THIS PROTEIN IS POSTULATED TO ACT BOTH AS TERMINAL ENERGY ACCEPTOR (BY ITS PHYCOBILIN-LIKE DOMAINS) AND AS A LINKER POLYPEPTIDE (BY ITS REPEATS AND ARMS) THAT STABILIZES THE PHYCOBILISOME CORE ARCHITECTURE.
CC CC -1 SUBUNIT: PHYCOBILISOMES OF THIS ORGANISM ARE COMPOSED OF A TWO CYLINDER CORE, FROM WHICH SIX RODS RADIATE. THE CORE IS MAINLY COMPOSED OF ALLOPHYCOCYANIN ALPHA AND BETA CHAINS, AND OF THREE MINOR COMPONENTS: THE ALLOPHYCOCYANIN ALPHA-B CHAIN, A 18.3 kDa PEPTIDE, AND THE ANCHOR POLYPEPTIDE LCM.
CC CC -1 SUBCELLULAR LOCATION: ANCHORS THE PHYCOBILISOME PERPENDICULARLY TO THE STROMAL SURFACE OF THE THYLAKOID MEMBRANE.
CC CC -1 SIMILARITY: THE REPEATED DOMAINS ARE SIMILAR TO THE N-TERMINAL REGIONS OF PHYCOCYANIN ROD LINKER POLYPEPTIDS.
CC CC -1 SIMILARITY: THE PHYCOBILIN-LIKE DOMAINS ARE SIMILAR TO PHYCOBILINS FROM VARIOUS SPECIES.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL, M20806, AAA24873.1; -.
DR DR PIR, A35088, A35088.
DR DR HSSP, P00318, IB33.
DR DR InterPro: IPRO01297; PBS_linker.poly.
```

DR InterPro: IPR001659; Phycobillosome.
DR Pfam: PF00427; Pbs_Linker_poi; 4.
DR Pfam: PF00502; Phycobillosome; 2.
KM Phycobillosome; Electron transport; Photosynthesis; Repeat.
FT INIT_MET 0
FT DOMAIN 17 75 PHYCOBILIN-LIKE 1.
FT DOMAIN 76 143 PHYCOBILIN-LIKE LOOP.
FT DOMAIN 144 236 PHYCOBILIN-LIKE 2.
FT DOMAIN 237 284 ARM 1 (SPACING SEQUENCE).
FT REPEAT 285 409 I.
FT DOMAIN 410 546 ARM 2 (SPACING SEQUENCE).
FT REPEAT 547 669 II.
FT DOMAIN 670 743 ARM 3 (SPACING SEQUENCE).
FT REPEAT 744 869 III.
FT DOMAIN 870 953 ARM 4 (SPACING SEQUENCE).
FT REPEAT 954 1079 IV.
SQ SEQUENCE 1079 AA; 120325 MW; 87FE38F232BFC82 CRC64;

Query Match 3.3%; Score 104.5; DB 1; Length 1079;
Best Local Similarity 21.3%; Pred. No. 8.1;
Matches 71; Conservative 47; Mismatches 80; Indels 135; Gaps 21;

OY 62 PFKQSN--SPAIYVKPKAKESNTKONLFORAALDAEGFLVSEKILHP-----L 113
DB 467 FKEKTRNPTSPA---PFSKDRILIN-----QGPQINSQVSNPGARGEF 509
OY 114 PRTADPSV---QIAG-----NFAPVNEQPYARNLPVYGKLPDSIKGYVVRNG 157
DB 510 PGLSLPKVFRILDLPQTIGTKAKAKASIKFESSSTQAV-----TKAATLQ-- 554
OY 158 ANPLHEPYTGHHFFDGDGM-VHAKVEHGSAS-----YACRFT 194
DB 555 -----VEGRDVGVGGRKQVGEIKLENGQLSVREFIRALAKSDVFRKTYWTSLVYCKAI 607
OY 135 QTRNEVQERQGRVYPPKAIIGELHGTGIALMLFYARAAGIVDPAGHTGVANAGLYE 254
DB 608 E---YIHRRLGLRPYGRG--EINKYFDIAAKQGFYA---VVD---AIIINS--VEY 650
OY 255 NCRLLAMSEDDLPVOVOTPNNG---DKTVG--RPDPGQLESTMIAHPKYDPEGELF 308
DB 651 SE---AFGEDVYPEREYLPSPGVALLRQLR-VGSIREVDVG-----KVQKQETPLF 696
OY 309 ALSDYVNSKPYLKYFRFSPDGTSPDVEIQLDQ 341
DB 697 VTLGTVT-----DTRTEPDICRIRNG 717

RESULT 8
SYFB_AQUAE STANDARD; PRT; 775 AA.
AC 067620;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
GN tRNA ligase beta chain) (PheRS).
OS PHE1 OR AO.1730.
OC Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.J., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus";
RL Nature 392:353-358(1998).
CC -I- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC diphosphate + L-phenylalanyl-tRNA(Phe).

CC -I- SUBUNIT: Tetramer of two alpha and two beta chains (by
CC SIMILARITY).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: BELONGS TO THE PHENYLALANYL-tRNA SYNTHETASE BETA
CC CHAIN FAMILY. SUBFAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000753; AAC07582.1; -.
DR HSRP: P27002; 1PYS.
KM Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KM Complete proteome.
SQ SEQUENCE 775 AA; 88484 MW; 61BB15FDE5B274C8 CRC64;

Query Match 3.3%; Score 103; DB 1; Length 775;
Best Local Similarity 20.7%; Pred. No. 6.5;
Matches 114; Conservative 67; Mismatches 159; Indels 212; Gaps 31;

OY 123 IAGNFAPVNEQPYARNLPVYGKLPDSIKGYVVRNGANPLHEPTGHHFDP----- 172
DB 46 VEGKVEVEGHEHPTKRLAV-----KVQYQEHIFIDVTVYDKSVRE 86
OY 173 GGMVHAKVEHGSASACRFTQTRNEVQERQGRVYPPKAIIGELHGTGIALMLFYAR 232
DB 87 GDGVI--VALPNAKVGNMG-----VTEREP-----DGVSKGILLSA 121
OY 233 AAAGIVDPAGHTGVANAGLYVRNGRLAMSEDDLP-----YQVOTPN-GD 277
DB 122 QELGLEEKSEG-----VLKIHDFKPGTDANEILGEGEKIIEIDITPNRQD 167
OY 278 LKTVGRFDDQLESTMIAPKVPD-----ESGELFALSYDVNSKPYLKYFRFSPDGTK 331
DB 168 MLSVGRVARD--LSAIFLPRKKRPEEPYEEETGEFFIEIDEDCKRY-----RGVY 216
OY 332 SPDVEIQLDQ-----TMKHDPAITENFVVPDQGVFKLPMIRIGSPV-VY 378
DB 217 IGVGEIK-ESPLYIKRLMQCGIKSINNVDITNVY-----MLRDGQPLHAF 262
OY 379 DNKVARFGILDKYADSSNIMKIDAPDCFCFLNMANMEPELDEVVYIGSCMTPPDSIF 438
DB 263 DLSKV-EGGIIVRSARKGKEIITLDG-----EREIDEDILVIADREKRLAVA 309
OY 439 N-----ES--DENLKSVLSEI-----RLNLKGTSTRRPIISNEDQVNLBA 478
DB 310 GVIIGLESIKENTKDILLESAYFNPFVRKASKKIGITESSYR-----FERAVDIE- 362
OY 479 GAVNRNMLGRKTKFA-YIALABPWPVKSGFAKAVDITLTTGFKHKLXGDNRYGGEPLFP-- 535
DB 363 -----RVDRADQYAVYLLIKHAGKV-----KVAVDYRE-KYRKRVYLPFG 404
OY 536 -----GEGGEDGECYILCFVHDEKTKWKSELDQIVANVSLFEVATYKLPSRVRYGFGHTG- 588
DB 405 KYIRVAGESYKNNE-----VKEIIDALEIPMEI-MRCGVLEVLP3-----HRSFD 448
OY 589 IGAD-DLAKQV 599
DB 449 IORDVDLIEIM 460

RESULT 9
CYAA_USTMA STANDARD; PRT; 2493 AA.
AC P49606;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenyl-
 DE cyclase).
 GN UAC1 OR REML.
 OS Ustilago maydis (Smut fungus).
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
 OX NCBI_Taxid=5270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=518;
 RX MEDLINE=95087882; PubMed=7995519;
 RA Gold S., Duncan G., Barrett K., Kronstad J.W.;
 RT "CAMP regulates morphogenesis in the fungal pathogen Ustilago
 RT maydis.";
 RL Genes Dev. 8:2805-2816(1994).
 CC -1- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR
 CC METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER,
 CC CAMP.
 CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
 CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.
 CC -1- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
 CC -1- SIMILARITY: CONTAINS A PP2C-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, L33918; AAA57469.1; -;
 DR InterPro: IPR001054; Guanylyl_cyclase.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003591; LRR_cyp.
 DR InterPro: IPR001932; PP2C_domain.
 DR Pfam: PF00211; guanylate_cyc; 1.
 DR Pfam: PF00560; LRR; 18.
 DR Pfam: PF00481; PP2C; 1.
 DR SMART; SM00044; CYCC; 1.
 DR SMART; SM00370; LRR; 6.
 DR SMART; SM00369; LRR_Typ; 2.
 DR SMART; SM00332; PP2C; 1.
 DR SMART; SM00314; RA; 1.
 DR PROSITE: PS50125; GUANYLATE_CYCLASES_2; 1.
 DR Lyase; Repeat; Leucine-rich repeat; CAMP synthesis; Magnesium.
 KM DOMAIN 759 763
 FT DOMAIN 888 897
 FT REPEAT 1084 1106
 FT REPEAT 1108 1131
 FT REPEAT 1132 1155
 FT REPEAT 1157 1178
 FT REPEAT 1179 1201
 FT REPEAT 1202 1225
 FT REPEAT 1227 1248
 FT REPEAT 1249 1271
 FT REPEAT 1273 1295
 FT REPEAT 1314 1337
 FT REPEAT 1338 1360
 FT REPEAT 1361 1384
 FT REPEAT 1386 1407
 FT REPEAT 1408 1430
 FT REPEAT 1432 1455
 FT REPEAT 1509 1529
 FT REPEAT 1533 1556
 FT REPEAT 1557 1580
 FT REPEAT 1581 1604
 FT REPEAT 1606 1628
 FT REPEAT 1633 1656
 FT REPEAT 1722 2001
 FT DOMAIN 2002 2493
 FT CATALYTIC.

SQ SEQUENCE 2493 AA; 271979 MW; 106A872C3C1C5BF8 CRC64;
 Query Match 3.2%; Score 102; DB 1; Length 2493;
 Best Local Similarity 21.7%; Pred. NO. 42;
 Matches 85; Conservative 44; Mismatches 136; Indels 126; Gaps 17;
 OY 4 FTATAVSGRWLGNT-----OPLSSOSSDLSYCSLPMASVYTKLVNLSALHTP 57
 DB 801 FKKTASLAGSRGTDSDVPLTALPLPGSKSVDEAAANKVVDLO-TNNLQSAALVVO- 858
 OY 58 PALHPPKQSSNSPAIVYKPKAKESNTKQNNLFQRAAAAALDAEGLVSH---EKHLPL 113
 DB 859 ---QSOQNHHQPSPVNVRTSRGAGAHM--PASAGASAAAAGKGLHRSKRHMNR 913
 OY 114 PRTAD-----PSYQIAGFAPVNEQPVARNLPYVGKLPDSIKGYVVRGANGPLHEPTG 167
 DB 914 PNTAGSVGATRPSTYTLTGSLSAEDD-----TSINGSIIRDG-HPLRSATA 959
 OY 168 -----HHFFDGDGMVAHVFEGHSASVACRFQTQNFVQERQGRVFPKAIGE 216
 DB 960 NNNMATGTLPRNH-----IKYKTIDGTFATLSCSLVSTANEVQ----- 998
 OY 217 LHGTGIARLMDFYRAAAGIVDPAHGTGVANAGLVYFNGRLAMSEDDLPRYOQ----- 271
 DB 999 ---FLARKSLTTESAAYRLF-----VRDKG---SERDLGIS--DKPSQLRRRLI 1041
 OY 272 ---TTPNGDLKTVGRDPFGQLESTMIAPKVPDESSELFAISDYVSKPYIKYRFSFD 328
 DB 1042 QAGYENDLEDEMGRDLS-----YLLREVFEPD 1070
 OY 329 GTRKSPDVE-IQDOPTMHDPATENFVVVP 358
 DB 1071 SVPTFDSISIGSEHTFQHLDSRLNEMVP 1101
 RESULT 10
 GLGS BRANA STANDARD; PRT; 520 AA.
 AC 09M462;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Glucose-1-phosphate adenyllyltransferase (small subunit, chloroplast
 DE precursor (EC 2.7.7.27) (ADP-glucose synthase) (ADP-glucose
 DE pyrophosphorylase) (AGPase B) (Alpha-D-glucose-1-phosphate adeny-
 DE ltransferase).
 GN AGP51.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_Taxid=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. Drakkar; TISSUE=Seed;
 RA Zawodny S., Martini N.;
 RT "Isolation and analysis of a cDNA clone encoding the small subunit of
 RT ADP-glucose pyrophosphorylase in the plastids of seeds and leaves of
 RT oilseed rape (Brassica napus).";
 RL Submitted (Jan-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS PROTEIN PLAYS A ROLE IN SYNTHESIS OF STARCH.
 CC IT CATALYSES THE SYNTHESIS OF THE ACTIVATED GLYCOSYL DONOR,
 CC ADP-GLUCOSE FROM G1C-1-P AND ATP
 CC -1- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
 CC diphosphate + ADP-glucose.
 CC -1- ENZYME REGULATION: ACTIVATED BY 3'PHOSPHOGLYCERATE, INHIBITED
 CC BY ORTHOPHOSPHATE. ALLOSTERIC REGULATION.
 CC -1- PATHWAY: STARCH BIOSYNTHESIS.
 CC -1- SUBUNIT: HETEROTETRAMER.
 CC -1- SUBCELLULAR LOCATION: Chloroplast (By similarity).
 CC -1- TISSUE SPECIFICITY: LEAVES.
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-

CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ271162; CAB89863.1; -.
DR InterPro: IPR001825; NTP transferase.
DR Pfam: PF00483; NTP transferase; 1.
DR PROSITE: PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE: PS00809; ADP_GLC_PYROPHOSPH_2; 1.
DR PROSITE: PS00810; ADP_GLC_PYROPHOSPH_3; 1.
DR Glycogen biosynthesis; Transferase; Nucleotidyltransferase;
KM Multigene family: Starch biosynthesis; Allosteric enzyme;
KM Chloroplast; Transit peptide.
FT TRANSIT 1 71 CHLOROPLAST (POTENTIAL)
FT CHAIN 72 520 GLUCOSE-1-PHOSPHATE ADENYLYLTRANSFERASE
FT SMALL SUBUNIT.
SQ SEQUENCE 520 AA; 57044 MW; D0EAF09706F3B6A7 CRC64;

Query Match 3.2%; Score 101.5; DB 1; Length 520;
Best Local Similarity 27.0%; Pred. No. 4.7;
Matches 43; Conservative 27; Mismatches 66; Indels 23; Gaps 7;

QY 1 MASTATAAAGRMVLCGHNTPPLSSQSSDLSSCPMASRYTRKLNVSALHTPAL 60
DB 1 MATMAALGSLKVPSSSSNHTRLSSSSORKTLSRSS-----SLGKELNPPQETIIS--- 53

QY 61 HFPKSSNSPAIVKPKA-KESNTKOMNLFQRAAALDAEGLVY---SHEKHLPL-PK 115
DB 54 NLPGNERRRPTISYSPKAVDSQNSQCTLDDPSRSVL---GIILGGAGTRILYPLRK 109

QY 116 TADPSVQIAGNFAPVNEQPVRRNLPVVGKLPDSIKGVYV 154
DB 110 RAKPAVPLGANYRLI-----DIPVSNCLSNISNISKIV 141

RESULT 11
SCRL_SCHPO STANDARD; PRT; 565 AA.
ID SCRL_SCHPO
AC 014335;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-binding protein scrl.
GN SCRL OR SPBC1D7.02C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCRL_TaxID-4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Wood V., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP IDENTIFICATION, AND FUNCTION.
RX MEDLINE-98204792; PubMed-9535817;
RX Tanaka N., Ohuchi N., Mukai Y., Osaka Y., Ohtani Y., Tabuchi M.,
RA Bhuiyan M.S.A., Fukui H., Harashima S., Takegawa K.;
RT "Isolation and characterization of an invertase and its repressor
RT genes from Schizosaccharomyces pombe.";
RL Biochem. Biophys. Res. Commun. 245:246-253(1998).
CC -1- FUNCTION: INVOLVED IN CARBON CATABOLITE REPRESSION. REPRESENTS
CC THE TRANSCRIPTION OF VARIOUS GENES INCLUDING THE INV1 GENE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE CREB/MIG GROUP OF C2H2-TYPE ZINC-

CC FINGERS PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z98270; CAB10978.1; -.
DR InterPro: IPR000822; ZnF-C2H2.
DR Pfam: PF00096; Zf-C2H2; 2.
DR PRINTS: PR00048; ZINCFTINGER.
DR SMART: SM00355; ZnF_C2H2; 2.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 2.
KM DNA-binding; Transcription regulation; Repressor; Zinc-finger;
KM Nuclear protein; Zinc; Metal-binding; Repeat.
FT ZN_FING 26 48 C2H2-TYPE.
FT ZN_FING 54 78 C2H2-TYPE.
FT DOMAIN 81 97 ALA-RICH.
FT DOMAIN 164 167 SER-RICH.
FT DOMAIN 375 378 POLY-GLN.
FT DOMAIN 375 378 POLY-SER.
SQ SEQUENCE 565 AA; 59713 MW; FE6CCCF7E54A8CDB CRC64;

Query Match 3.2%; Score 101.5; DB 1; Length 565;
Best Local Similarity 21.8%; Pred. No. 5.3;
Matches 74; Conservative 38; Mismatches 147; Indels 81; Gaps 15;

QY 6 ATAAVSGRMVLCGHNTPPLSSQSSDLSSCPMASRYTRKLNVSALHTPAL 60
DB 147 ATAAVSMYV--PHHSASVQOQATFVNSGQPHNLPQAQAPRTIVGIDPLHTTQNGTTI 204

QY 61 HFPKSSNSPAIVKPKAKESNTKOMNLFQRAAALDAEGLVY---SHEKHLPL-PK 120
DB 205 HVTGTPGCAVAVSQRSERDRLSSNMQLASAAANQLDAA-----PRIT-PT 250

QY 121 VQIAGNFAPVNEQPVRRNLPVVGKLPDSIKGVYVRNCGANPLH-EPVYGHFFDDGAVH 178
DB 251 KSSGVNLMPLSNAPSPPPKQMNVYGLSPSS-----SNTSPHLASVPNRKLTSSSTGS 303

QY 179 AVKEFGSASVACRFTQNNRFQERQLGRPVY---KAIGELGHTGIARLMFYARAAA 235
DB 304 FTKYING-----SNLSYNSMQTPYLPKSKNSSTSLHSMYGVGT----- 344

QY 236 GIVDPAHGTVANAGLVYFNGLAMSEDDLPYQVQITPNGLKTVGRFPEDGLESTWI 295
DB 345 -----TAYAPQSLRYAH-----YNYLPSRPSVSN-----FDDSSSDPF 380

QY 296 AHPKVDPESGELFALSYDVVSKPYLKYFRFSPDGKRSPDV 335
DB 381 AHFRYQRSRPVPCS-TAPSSPTSTSTSP-----TPDV 415

RESULT 12
GIGT_VICFA STANDARD; PRT; 512 AA.
ID GIGT_VICFA
AC P52417;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase small subunit 2, chloroplast
DE precursor (EC 2.7.7.27) (ADP-glucose synthase) (ADP-glucose
DE pyrophosphorylase) (AGPase B) (Alpna-D-glucose-1-phosphate adenylyl
DE transferase).
GN AGPP.
OS *Vicia faba* (broad bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids I; Fabales; Fabaceae; Papilionoideae; Viciae; *Vicia*.

OX NCBI_TaxID=3906;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, FRIBO; TISSUE=Cotyledon;
 RX MEDLINE=95216601; PubMed=7766042;
 RA Weber H., Helm U., Borisjuk L., Mobus U.;
 RT "Cell-specific, coordinate expression of two ADP-glucose
 pyrophosphorylase genes in relation to starch biosynthesis during
 seed development of *Vicia faba* L.";
 RL Planta 195;352-361(1995).
 CC
 CC -1- FUNCTION: THIS PROTEIN PLAYS A ROLE IN SYNTHESIS OF STARCH.
 IT CATALYZES THE SYNTHESIS OF THE ACTIVATED GLYCOSYL DONOR,
 ADP-GLUCOSE FROM GLC-1-P AND ATP.
 CC
 CC -1- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
 diphosphate + ADP-glucose.
 CC
 CC -1- ENZYME REGULATION: ACTIVATED BY 3'PHOSPHOGLYCERATE, INHIBITED
 BY ORTHOPHOSPHATE. ALLOSTERIC REGULATION.
 CC
 CC -1- PATHWAY: STARCH BIOSYNTHESIS.
 CC
 CC -1- SUBUNIT: HETEROTETRAMER.
 CC
 CC -1- SUBCELLULAR LOCATION: Chloroplast (By similarity).
 -1- TISSUE SPECIFICITY: LEAVES AND SEEDS.
 CC
 CC -1- DEVELOPMENTAL STAGE: IT IS PRESENT IN YOUNG COTYLEDONS AT 14 DAYS
 AFTER FERTILIZATION (DAF) WHEN CELLS ARE STILL RAPIDLY DIVIDING.
 CC
 CC LEVELS STEADILY ACCUMULATE UNTIL THE END OF THE CELL EXPANSION
 PHASE (35-40 DAF) AND WITH THE BEGINNING OF THE SEEDS DESICCATION
 PHASE AT 50 DAF, THE LEVELS DECREASE TO VERY LOW LEVELS.
 CC
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
 PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X76941; CAA54260.1; -;
 DR InterPro: IPR001825; NTP_transferase.
 DR Pfam; PF00483; NTP_transferase; 1.
 DR PROSITE; PS00808; ADP_GLC_PYROPHOSPH_1; 1.
 DR PROSITE; PS00809; ADP_GLC_PYROPHOSPH_2; 1.
 DR PROSITE; PS00810; ADP_GLC_PYROPHOSPH_3; 1.
 KW Glycogen biosynthesis; transferase; Nucleotidyltransferase;
 KW Multisubstrate; Starch biosynthesis; Allosteric enzyme;
 KW Chloroplast; Transl. peptide.
 FT TRANSIT 1 63 CHLOROPLAST (POTENTIAL).
 FT TRANSIT 64 512 GLUCOSE-1-PHOSPHATE ADENYLYLTRANSFERASE
 FT CHAIN 1 63 SMALL SUBUNIT 2.
 FT FT
 FT SEQUENCE 512 AA; 56059 MW; 060647F683526EDC CRC64;
 SQ
 Query Match 3.2%; Score 101; DB 1; Length 512;
 Best Local Similarity 21.0%; Pred. No. 5; Indels 206; Gaps 30;
 Matches 118; Conservative 68; Mismatches 170;

Db 222 ESDADITVALMDERARA-----FGLMKIDEGRIIEFSEN----- 259
 QY 274 PNGD-LK-----YGRFPPQOLESTMTA-----HPKYD----- 301
 Db 260 PGEQKAKKAVDTLLGLDDRAKEMPHYIASMGIVYVSKHVMILDRKPGANDFGSEV 319
 QY 302 -PESGSL-----PALSYD-----VSKP-----YLKYPFSPDGKSPV 335
 Db 320 IFGATELGRVQAYLYLDGWEDIGTEAFYNNALGITKPVDPFSFYDRSSPLYTOP--- 376
 QY 336 ETLDDPTMMDFAITENFY---VVPDOV---VEKLPENI----- 370
 Db 377 -RYLPKSKMLADITDSVIGEGCYIKNCKIHSHVGLNSCSISEGAIIDTLMGADYVE 434
 QY 371 -----RGSPVYVDKKNVAREGLIDKTAEDSSNKKWJADPDCFCFHLNMAEPE 420
 Db 435 TDADRRLAANKGVPVIGICKNSHIRALIDKNARIGDVKIINS-----NVQEAARETE 489
 QY 421 -----TDEVVYVSGCWTPPDSI 437
 Db 490 GFYFISGIVTVIKDLISGTV 511
 RESULT 13
 ACVS_EMENT STANDARD; PRT: 3770 AA.
 ID ACVS_EMENT
 AC P27742.
 DT 01-AUG-1992 (rel. 23, Created)
 DR 01-AUG-1992 (rel. 23, Last sequence update)
 DR 16-OCT-2001 (rel. 40, Last annotation update)
 DE Delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase
 DE (EC 6.) (ACV synthetase) (ACVS).
 GN ACVA.
 OS *Emmericella nidulans* (*Aspergillus nidulans*).
 OC Euryarchaeota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiiales; Trichocomaceae; *Emmericella*.
 OX NCBI_TaxID=5072;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=G191;
 RX MEDLINE=91286299; PubMed=2061333;
 RA Maccabe A.P., van Liempt H., Pallissa H., Unkles S.E., Riach M.B.R.,
 RA Pfeiffer E., von Doehren H., Kinghorn J.R.;
 RT "Delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from
Aspergillus nidulans. Molecular characterization of the acva gene
 encoding the first enzyme of the penicillin biosynthetic pathway.",
 RT J. Biol. Chem. 266:12646-12654(1991).
 RL J. Biol. Chem. 266:12646-12654(1991).
 CC -1- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE
 ACV ARE ACTIVATED AS AMINOADYL-ADENYLATES WITH PEPTIDE BONDS
 FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
 INTERMEDIATES.
 CC
 CC -1- COPACATOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANETHEINES
 (POTENTIAL).
 CC
 CC -1- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND
 CEPHALOSPORIN.
 CC
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.
 CC
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 FAMILY.
 CC
 CC -1- SIMILARITY: CONTAINS 3 ACTYL CARRIER DOMAINS.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X54853; CAA38631.1; -;
 DR PIR; A40889; A40889.
 DR HSSP; P14687; IAMU.
 DR InterPro: IPR000873; AMP-bind.

DR InterPro: IPR001242; DUF4.
DR InterPro: IPR000379; Est_lip_thioest_actsite.
DR InterPro: IPR003880; Phosphopant_attach.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00501; AMP-binding; 3.
DR Pfam: PF00668; Condensation; 3.
DR Pfam: PF00550; PP-binding; 3.
DR Pfam: PF00975; Thioesterase; 1.
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 3.
DR PROSITE: PS00455; AMP_BINDING; 3.
DR PROSITE: PS50075; ACP_DOMAIN; 3.
DR Liasse: Antibiotic biosynthesis; Multifunctional enzyme;
Repeat; Phosphopantetheine.
FT REPEAT 321 910 DOMAIN 1 (ADIPATE-ACTIVATING).
FT REPEAT 1413 1993 DOMAIN 2 (CYSTEINE-ACTIVATING).
FT REPEAT 2494 3078 DOMAIN 3 (VALINE-ACTIVATING).
FT DOMAIN 850 919 ACYL CARRIER (ACP) 1.
FT DOMAIN 1929 2002 ACYL CARRIER (ACP) 2.
FT DOMAIN 3020 3087 ACYL CARRIER (ACP) 3.
FT BINDING 882 882 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 1965 1965 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3050 3050 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT ACI_SITE 3623 3623 THIOESTERASE (BY SIMILARITY).
SQ SEQUENCE 3770 AA; 422448 MW; CB66BD232A58CB0 CRC64;

Query Match 3.2%; Score 100; DB 1; Length 3770;
Best Local Similarity 20.9%; Pred. No. 1.le+02;
Matches 97; Conservative 67; Mismatches 162; Indels 138; Gaps 27;

QY 12 GMLGNTPTPLSSSSDLSYCSLPMASRYTRK-----LNVSAHLTPPAL 60
DB 1544 GMSWSN-----PAPSTSTDLAY-----IYTSGTGPKQVMEHHGVNVLQSLSTFEL 1596
QY 61 -----HPKQSS-----NSPAIVVKPKAKESNTKQMLFQRAAAALDA 99
DB 1597 RDTDEVILSFSNVYFDHVEQMTDALINGQTLVMLNDAMRSDKER--LYQLETNRVTV 1654
QY 100 AGCFVLSHEKHLRLPKTADPSVQIAGNFAPVNEQPVRR-----NLVYVKRLPDSTIG 151
DB 1655 LSG-----TPSYISMTEFSRFRKH--LRRVDVCGEAFSQVPEQDIRTFQ 1698
QY 152 VYVRNGANPLHEPVYTHNEFFDGDGMVHAKVFEGHSASYACRFQTNRFVQERQLGR-PVF 210
DB 1699 LII-NGYGTETISITTHK-----RLVPRPERRTDKSIGQIGNSTSYLNDMKNRPV 1750
QY 211 PRAIGELH-GHTGIRALMLFTAR-----AAAGIVDPAHGTVANNAGLVYNGRLLANSEDD 265
DB 1751 -GAVGELYLGEGVAR--GYNHRPEVTAERFLRNPFOJIDSEKONGR--NSRLYRTGD-- 1802
QY 266 LFYVOYIPL--NGDJKTVGRFDFD-----GOLSTMTAHP-----KYDPES 304
DB 1803 ---LYRMLPGSNGEIEYIGRNDFOYKINGRLTEIGETIAVMSHPDIKOSVVIANSKGKG 1859
QY 305 GELFALSTDVYVSKPYLKYFRSPDGTSPDVEIQIDQPTMMHDFATTENV-----VVP 358
DB 1860 DQKFLVGYFVASSP-----LSPGA-----IRRMQSRRLGVMYPPSFIFISSLPVPR 1906
QY 359 DQGVVFK-LPEM-IRGGSFVYVDKKNVAF-----GIIDKTA 394
DB 1907 SGKLDTKALPTAEKGMANVLAPRNEIESILGISAGLIDISAO 1950

RESULT 14
PHYC_BACSU
AC 031097; STANDARD; PRT; 383 AA.
DT 30-May-2000 (Rel. 39, Created)
DT 30-May-2000 (Rel. 39, Last sequence update)
DT 30-May-2000 (Rel. 39, Last annotation update)
DE 3-phytase precursor (EC 3.1.3.8) (Phytate 3-phosphatase) (MTO-inositol-hexaphosphate 3-phosphohydrolase).

GN PHYC OR PHYB13.
OS Bacillus subtilis.
OC Bacteria: Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1423;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-VTT E-68013;
RA MEDLINE-98268943; PubMed=9603817;
RX Kervuo J., Laureus M., Nurminen P., Kalkkinen N., Aapalahti J.,
RT "Isolation, characterization, molecular gene cloning, and sequencing
of a novel phytase from Bacillus subtilis."
RT Appl. Environ. Microbiol. 64:2079-2085(1998).
CC -FUNCTION: CATALYZES THE HYDROLYSIS OF INORGANIC ORTHOPHOSPHATE
FROM PHYTATE. ONLY PHYTATE, ADP, AND ATP WERE HYDROLYZED (100, 75,
AND 50% OF THE RELATIVE ACTIVITY, RESPECTIVELY). HAS MAXIMAL
ACTIVITY AT PH 7 AND 55 DEGREES CELSIUS.
CC -1 CATALYTIC ACTIVITY: Myo-inositol hexakisphosphate + H(2)O = D-myo-
inositol 1,2,4,5,6-pentakisphosphate + phosphate.
CC -1 COPACITOR: ISOLATED ENZYME REQUIRED CALCIUM FOR ITS ACTIVITY AND/OR
STABILITY.
CC -1 SUBCELLULAR LOCATION: Secreted.
CC -1 INDUCTION: BY PHYTATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF029053; AAC31775.1; -.
DR HSSP: 066037; 2PRO.
DR InterPro: IPR003431; Phytase.
DR Pfam: PF02333; Phytase; 1.
KW Hydrolase; signal.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 29
FT CHAIN 30 383 3-PHYTASE.
SQ SEQUENCE 383 AA; 41922 MW; E9BEC24A48EB9CA CRC64;

Query Match 3.2%; Score 99.5; DB 1; Length 383;
Best Local Similarity 19.3%; Pred. No. 4.2; Mismatches 151; Indels 147; Gaps 22;

QY 18 NHTQPLSSSSQSSDLSYCSLPMASRYTRKLN-----VSSALHTPPALHPKQSSNSPA 71
DB 2 NNSKTLTLTAAGIMLTGCA--VSSQAKHKLSDPYHFTVNAAEETPR-VDTAGDAADDP 58
QY 72 IYVVKPKAKESNTKQMLFQRAAAALDAAGFLVSHKHLRLPKTADPSVQIAGNFAPV 131
DB 59 IWDPRTPQ-NSR-----LTTNKKSGLVVYSLDGKMLSHYNNGKLN 99
QY 132 EQPVRRNLVYVKLDPDSIKGVYVRNGANPLHEPVYTHNEFFDGDGMVHAKVFEGHSASYAC 191
DB 100 NDIARDPFLNCKKVDIAAASNRSECKNIE-----IYALDGKNGTL----- 141
QY 192 RFTQTNRFVQERQLGRVPR--KAIGELHGTGIRALMLFYARAAGIVDPAHGTVANA 249
DB 142 -----OSMEDPRDPIATAIINEVGYT-----LYHSQK----- 168
QY 250 GLVYFNGRLAM--SEDL-PLYOYITPNSGL--KTVGRFPDQGLESTMTAHPKVDE 303
DB 169 -----TGKTYAANTGKEGFEBOETELADKNGYISGKVAFAKNSOTEGM-----AADE 218
QY 304 GELFALSTDVYVSKPYLKYFRSPDGTSPDVEIQIDQPTMMHDFATTENV 355
DB 219 YGRLLYAEDEA-----IMKFSAPRDOGSNGYIDRADGNHLRLDGLGTLTYAAADQKGYL 274
QY 356 VVPDQGVVFKLPEMIRGGSFVYVDKKNVAF-----KKK-VARQSIDKYEDSSNIMWIDAPDCFCFH 411

Db 275 MASSQ-----GNSSYAIVDRQKNKYVADFRTDGPETDGTG----- 311
QY 412 LMANEPEETDEVVIGSCWTP--PDSIFNEPD-ENL 445
Db 312 -----DIDGIDVLGFLGPEYPRGIFVADGENI 340

RESULT 15
ATL_STRAU ID ATL_STRAU STANDARD: PRT: 1256 AA.
AC P52081;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bifunctional autolysin precursor (includes: N-acetylmuramoyl-L-alanine
DE amidase (EC 3.5.1.28); Mannosyl-glycoprotein endo-beta-N-
DE acetylglucosamidase (EC 3.2.1.96)).
GN ATL.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
RC STRAIN-RN450;
RX MEDLINE=95116542; PubMed=7816834;
RA Oshida T., Sugai M., Komatsuzawa H., Hong Y.-M., Suglnaka H.,
RA Tomasz A.;
RT "A. Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
RT alanine amidase domain and an endo-beta-N-acetylglucosaminidase
RT domain: cloning, sequence analysis, and characterization.";
RT Proc. Natl. Acad. Sci. U.S.A. 92:285-289(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MCTC 8325-4;
RL Foster S.J.;
RL Submitted (Apr-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLCHITOSINYL UNIT IN
CC HIGH-MANNOSE GLYCOPROTEINS AND GLYCOPROTEINS CONTAINING THE
CC -(MAN)5(GLCNAc)2-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
CC RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
CC OLIGOSACCHARIDE IS RELEASED INTACT.
CC -1- CATALYTIC ACTIVITY: Hydrolyses the link between N-acetylmuramoyl
CC residues and L-amino acid residues in certain bacterial cell-wall
CC glycopeptides.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
CC EXTRACELLULAR LYTIC ENZYMES.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
CC ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
CC GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; DI7366; BAA04185.1; -
DR EMBL; LA1499; AAA99982.1; -
DR InterPro: IPR002502; Amidase_2.
DR InterPro: IPR002901; Amidase_4.
DR Pfam: PF01510; Amidase_2; 1.
DR Pfam: PF01832; Amidase_4; 1.
DR SMART; SM00047; LY22; 1.
DR Cell wall; Hydrolase; Signal; Multifunctional enzyme; Repeat.
FT CHAIN 1 29
FT SIGNAL 1 29
FT CHAIN 30 1256
FT DOMAIN 199 775
FT DOMAIN 776 1256
FT BIFUNCTIONAL AUTOLYSIN.
FT N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
FT ENDO-BETA-N-ACETYLGLUCOSAMIDASE.

FT REPEAT 425 589 1.
FT REPEAT 596 758 2.
FT REPEAT 770 932 3.
SQ SEQUENCE 1256 AA; 137384 MW; 2BB76CAA292FDD20 CRC64;

Query Match 3.2%; Score 99.5; DB 1; Length 1256;
Best Local Similarity 19.3%; Pred. No. 24;
Matches 89; Conservative 59; Mismatches 164; Indels 149; Gaps 20;

QY 7 TAAVSGRWLGG-----NHQDP-----PLSSOSSDLSYC-----SSLPMASRY 44
Db 98 TAAVNDGDTFRANOSATNTNTPQVAKSSTTAAPKTNVTWTAAGSLVDEDEDNSNQINPEL 157
QY 45 FRKLTVSSALHTPPALHPKROSSNSPAIVYKPAKESNTRKQMLFQRAAAALDAAEGLF 104
Db 158 IKSAAKPAALLETQYKTAAPKAATTS-----APRAKTEALTPKVTTFSSASAPRSVAA----- 208
QY 105 VSHEKLHPLEKTPADPSVOIAGNPAVNEQVRRNRPV-----VGKL 145
Db 209 -----TPKTSLPKYPKQVNV--SSINDYICKNNLAKPKIEDYTSYPPKYAYRNGVGR- 258
QY 146 PDSIKGYVYRANGANPLHEPYTG-----HHFPDGGWYHAAYKFEHGSASYAC 191
Db 259 PE---GIYVHDTAND--RSTINGEISYMKNNYQNAFYHAFVDDGRILLETAPTDYLSWGVA 314
QY 192 RFTQTRRFVQEROLGRPVFPKA-IGELHGHGTIARLMLFYARAAGIVDPAHGTGVANAG 250
Db 315 -----VGNPRFTNVEIVTHDYASPARSMNNTADYAA-----TQ 348
QY 251 LVYFNGRLAMSBDLLPYQVQITPNGDLKT--VGRF-----DFDQLESTMIAPHVY 300
Db 349 LQYYGLK-----PDSAEYDNGTVMTHYAVSKYLGCTGDHADPHGYLRSHNYSYDOL 399
QY 301 DPESGELFALSYDVVSKPYL-KYFRSPDGTSPDVIEQLDPTMHHDRATFNFPVVD 359
Db 400 -----YDLINKEKYLIMKGVAPMCTOSTTPTPTTPSKPT-----TPS 435
QY 360 QQVVPKLPKEMIRGSPVVDKKNVAFRGILDKKVAEDSSNIK 400
Db 436 KPSTGKLTVAANNQVAGQIKPTNSGLTYTYDRIKGAATNEVQ 476

Search completed: July 19, 2002, 10:16:06
Job time: 233 sec

Db 277 DCFESNEIMGVWLHIADKKRKYLNKKYTSSEFNT-----FHHINTYEDNEF--L 324
Qy 425 VVIGSCMPDPSIFNESD-ENLKSYLEIRLNKGTGSTRPRLISNEDQOVLNAGMVR 483
Db 325 IYDLCQWKGFEFVYNYLYLANLRNMEEVK---KNARKAPQPEVRSVPLNIDKRDYK 381
Qy 484 NMLGRKTKFAYIAL-----AEPMPKVSQ-----FAKVDLTT--GEVKKHLYG----- 523
Db 382 NLVTLPTATATATLNSDETITMLEPEVLFGSPROAFEPQINQYKGKPYTYAAGLGNH 441
Qy 524 -----DNRYGGEPLFLP-GEGGEDEGYILCFVHDEKTK--SE 559
Db 442 EYVDRCLKLVNKTETWQOEPDPSPEPIFVSHPDALBEDDGVLSVVSFGAGCKPAY 501
Qy 560 LQIVNAVSLVEEATVKLPSPVPGFHGT 588
Db 502 LLILNAKDLSEVARAEVEINIPVTFHGLF 530

RESULT 2
US-08-488-305A-6
; Sequence 6, Application US/08488305A
; Patent No. 5619772
; GENERAL INFORMATION:
; APPLICANT: B VIK, Claes Olof, Eriksson, ULf, Peterson, Per A.
; TITLE OF INVENTION: Isolated protein Receptors, Antibodies which
; TITLE OF INVENTION: blind thereto, Nucleic Acid Sequence Coding
; Patent No. 5619772
; TITLE OF INVENTION: Theretofor, And Uses Thereof
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felte & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488-305A
; FILING DATE: 7-JUNE-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohli, Vinset
; REGISTRATION NUMBER: 37,003
; REFERENCE/DOCKET NUMBER: LUD 5280.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-488-305A-6

Query Match 6.6%; Score 209; DB 1; Length 533;
Best Local Similarity 20.7%; Pred. No. 6.1e-13;
Matches 119; Conservative 99; Mismatches 222; Indels 134; Gaps 29;

Qy 105 VSEKLIHPLPKRADPSVOIAGNFAPVNRPNLP-----VVGKLPDSIKGYVVRNGANP 160
Db 1 MSSQVEHP-----AGGYKFLFETVEELSSPLTAHVGRPLMLWTGSLRCGPGL 49
Qy 161 LH---EPVYGHFFDGDGVNVAHVKFEGSASACFTOTNRVOE-----ROLGRVY 209
Db 50 FEVGSSEPF--YHLPDQALILHKFKEGHVYTHRRFIRIDAVYRAMTEKRIYITEFGICA 107

Qy 210 FPKAIGELHGHTGIARLMLFARAAAAGIVPAHGTGVANAGLVFNGRLAMSED----- 264
Db 108 FPDPCNKI-----FSRFFSYF-----RGVEVTDMALV-----NLYPGEDYACT 147
Qy 265 DLPYQOITPNNDLTKVGFDPDGOLE-STYLAHFKVNPES-----GELFALSDYV 314
Db 148 ETNFTITKVNPE-TLETIKQVDCNYYSVNGATAPHIENDGTVYVNIQNGCFGNFSIAYNI 206
Qy 315 VSKPYLKYFRFSPDGRKSPDVEIOLD-----QPTMHDEALITENFVVPDQOVFKLPM 369
Db 207 VAIPLPADKDEP--ISKSEIYVQPCSDRFKPSYVHSGLPNTIYVETVYKINKLRF 264
Qy 370 IRGGSFVYDKNKVAE-----GLDKYAEEDSNIKIDAPDCCFLMANAMEEP 419
Db 265 LSSWS--LWGANYMDCFESENETMGVWLHIADKKRKYLNKKYTSRP-FULFHHINTYEDH 321
Qy 420 ETDEVVYIGSCMTPDPSINESD-ENLKSYLEIRLNKGTGSTRPRLISNEDQOVLNLEA 478
Db 322 EF--LIVDLCQWKGFEFVYNYLYLANLRNMEEVK---KNARKAPQPEVRRYVPLNDK 376
Qy 479 GMYNRNM--LGRKTKFAYIALAE-PKV--SG-----FAKVDLTT--GEVKKHLYG 523
Db 377 ADTGKRLVTLPTATATALCSDETITMLEPEVLFGSPROAFEPQINQYKGKPYTYAYG 436
Qy 524 -----DNRYGGEPLFLP-GEGGEDEGYILCFVHDEKTK 556
Db 437 LGLNHFPDRCLKLVNKTETWQOEPDPSPEPIFVSHPDALBEDDGVLSVVSFGAG 496
Qy 557 K--SELQIVNAVSLVEEATVKLPSPVPGFHGT 588
Db 497 QRPAYLLILNAKDLSEVARAEVEINIPVTFHGLF 530

RESULT 3
US-08-928-361B-30
; Sequence 30, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, CARYN, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928-361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, HANA
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1043 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-30

Query Match 3.9%; Score 123.5; DB 3; Length 1043;
Best Local Similarity 20.6%; Pred. No. 0.0018;
Matches 125; Conservative 85; Mismatches 194; Indels 203; Gaps 35;

QY 5 TATAVSGRWLGNGHTOPPLSSOSSDLSYCSSTL-PMASRYTRK---LNVSSALHTPP 58
DB 153 TIAGIVSG-----ISASESL-LSQKSALIDPATNMVYGEFGLNPAATGVMIPG 200
QY 59 ALHFPKQSSNSPAI---VVKPKAKESNTKQML-----FQRAAAALD 98
DB 201 FLGPEQOTQSPFEIDGGIIPPEVAANAADKFKLSIPSPVESIPEKDQKIDISIELMYD 260
QY 99 AAEGLVSHKRLHPLPKTADPSVQIAGNFAPVNEQPVARNLPVVGKLPDSIKGYVVRNGA 158
DB 261 IESGRLIQVSKRPPIGS-----IAGDLNPIMKPTPTQD-SYTGKPIPTTGL----- 307
QY 159 NPLHEPVTHHFFPDGDGVAHAKFEHGSASY--ACRFTQTNKFEVQEROLGRVPFKAIGE 216
DB 308 -PFNPP-TGH-----LINPTNNTMDSFAGAYKAVSNGIKTDNYGLPV-----GE 353
QY 217 LHGHTGIARMLFYARAAGIYDPAHGTGVAN--AGLYVFNGRL-LAMSEDDLPTVOYIT 273
DB 354 ITGLPKDPSDIPFNSTTGEIVDPSTGKPINNSTAGIV--SGKPGLPRIEDE----- 403
QY 274 PNGDLKTYGRD-----FDQLTESTMTAHPKVDPSGELFALSYDVVSKPYLKYFRFSP 327
DB 404 -NGNL-----FDPSTNLPIDGN-----NQLVNETNSTYSGSYSGTKP-----KP 443
QY 328 -----DGTSPDVEI--OLDQ-----PTMMHDFATENFNVVPPDOQVVF 364
DB 444 GIPVNGGCVVPPEAKDDAKDKDGLIYPTNSINKDPTVNTQYNTGNTGNIINPETGKY- 502
QY 365 KLEPMIRG-----GSPVYVDKKNVAFRGILDKYAEDSSNIKWDADPCFCHILMANWE 417
DB 503 -IPGSLPGSLNYPSPNTPOQTDE-----ITGKPVDTVTGLPY----- 538
QY 418 EPEIDVAVVIGSCMTPPDSIFNESDENLKSYLEIRLNLTGSESTRRPLISNEDQVNL 477
DB 539 DPTGEIIDPATKLPPIGSVAGD-----ELITEV-LNITDEVVGLPI-----DLE 583
QY 478 AGMVNRNMLGRKTKFAYIALAEPMPKVSGEFAKV--DLTGEVKKHLYGDNRYGGEPLFL 534
DB 584 TGL-----PRDPVSGLPOLPMTGLVDPSPNKKPIRGS--HSG---FI 619
QY 535 PGEKGEE 541
DB 620 NCTSGEQ 626

RESULT 4
US-08-700-651-5

Sequence 5, Application US/08700651B

Patent No. 6013682

GENERAL INFORMATION:

APPLICANT: PETERSEN, CAROLYN

APPLICANT: LEECH, JAMES

APPLICANT: NELSON, RICHARD, C.

APPLICANT: GUT, JIRI

TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS

TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum

TITLE OF INVENTION: INFECTIONS

FILE REFERENCE: 480.19-4 (HV)

CURRENT APPLICATION NUMBER: US/08/700.651B

CURRENT FILING DATE: 1997-08-14

EARLIER APPLICATION NUMBER: 08/415,751

EARLIER FILING DATE: 1995-04-03

NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 1721
TYPE: PRT
ORGANISM: Cryptosporidium parvum
US-08-700-651-5

Query Match 3.9%; Score 123.5; DB 3; Length 1721;
Best Local Similarity 20.6%; Pred. No. 0.0044;
Matches 125; Conservative 85; Mismatches 194; Indels 203; Gaps 35;

QY 5 TATAVSGRWLGNGHTOPPLSSOSSDLSYCSSTL-PMASRYTRK---LNVSSALHTPP 58
DB 831 TIAGIVSG-----ISASESL-LSQKSALIDPATNMVYGEFGLNPAATGVMIPG 878
QY 59 ALHFPKQSSNSPAI---VVKPKAKESNTKQML-----FQRAAAALD 98
DB 879 FLGPEQOTQSPFEIDGGIIPPEVAANAADKFKLSIPSPVESIPEKDQKIDISIELMYD 938
QY 99 AAEGLVSHKRLHPLPKTADPSVQIAGNFAPVNEQPVARNLPVVGKLPDSIKGYVVRNGA 158
DB 939 IESGRLIQVSKRPPIGS-----IAGDLNPIMKPTPTQD-SYTGKPIPTTGL----- 985
QY 159 NPLHEPVTHHFFPDGDGVAHAKFEHGSASY--ACRFTQTNKFEVQEROLGRVPFKAIGE 216
DB 966 -PFNPP-TGH-----LINPTNNTMDSFAGAYKAVSNGIKTDNYGLPV-----GE 1031
QY 217 LHGHTGIARMLFYARAAGIYDPAHGTGVAN--AGLYVFNGRL-LAMSEDDLPTVOYIT 273
DB 1032 ITGLPKDPSDIPFNSTTGEIVDPSTGKPINNSTAGIV--SGKPGLPRIEDE----- 1081
QY 274 PNGDLKTYGRD-----FDQLTESTMTAHPKVDPSGELFALSYDVVSKPYLKYFRFSP 327
DB 1082 -NGNL-----FDPSTNLPIDGN-----NQLVNETNSTYSGSYSGTKP-----KP 1121
QY 328 -----DGTSPDVEI--OLDQ-----PTMMHDFATENFNVVPPDOQVVF 364
DB 1122 GIPVNGGCVVPPEAKDDAKDKDGLIYPTNSINKDPTVNTQYNTGNTGNIINPETGKY- 1180
QY 365 KLEPMIRG-----GSPVYVDKKNVAFRGILDKYAEDSSNIKWDADPCFCHILMANWE 417
DB 1181 -IPGSLPGSLNYPSPNTPOQTDE-----ITGKPVDTVTGLPY----- 1216
QY 418 EPEIDVAVVIGSCMTPPDSIFNESDENLKSYLEIRLNLTGSESTRRPLISNEDQVNL 477
DB 1217 DPTGEIIDPATKLPPIGSVAGD-----ELITEV-LNITDEVVGLPI-----DLE 1261
QY 478 AGMVNRNMLGRKTKFAYIALAEPMPKVSGEFAKV--DLTGEVKKHLYGDNRYGGEPLFL 534
DB 1262 TGL-----PRDPVSGLPOLPMTGLVDPSPNKKPIRGS--HSG---FI 1297
QY 535 PGEKGEE 541
DB 1298 NCTSGEQ 1304

RESULT 5
US-08-928-361B-6

Sequence 6, Application US/08928361B

Patent No. 6071518

GENERAL INFORMATION:

APPLICANT: Petersen, Carolyn

TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,

TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS

TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM

TITLE OF INVENTION: SPECIES INFECTIONS

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: PETERS, VERNY, JONES & BIRSA

STREET: 385 Sherman Avenue, Suite 6

CITY: Palo Alto

STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vermy, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1721 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-6

Query Match 3.9%; Score 123.5; DB 3; Length 1721;
Best Local Similarity 20.6%; Pred. No. 0.0044;
Matches 125; Conservative 85; Mismatches 194; Indels 203; Gaps 35;

QY 5 TPTAAVSGHMLGNHTQPLSSQSSDLSYCSL--PMASTRK---LNVSSALHTPP 58
DB 831 TTAGTIVS-----ISASESL-LSQKSLIDPATNMVYGERGLINPATGVIMIG 878
QY 59 ALHFPRKSSNSPAL---VVKPKAKESNTKQNL-----FQRAAALD 98
DB 879 FLGPSEQOTPFSPEDDGLIIPPEVAANADKFKLSIPSPVESIPEKDKIDISIELMYD 938
QY 99 AAEGLVSHKHLPLPKTADPSVOIAGNFAPVNEQVRRNLPPVYGLPDSIKGVYRNGA 158
DB 939 IESGRILQGVSKRIPGS-----IAGDLNPKIPTQTD-SYTGKRPIDPTTGL----- 985
QY 159 NPLHEPVYGHFFPDGDGVAHAKFEHGSASY--ACRFOTNRFOEROLGRPVFKPAIGE 216
DB 986 -PFNRP-TGH-----LINPTNNNTMDSFAGAYKVAVNSGIKTDNVYGLPV-----GE 1031
QY 217 LIGHGTIARLMLFYAARAAGIYDPAHGTGVAN--AGLYVFNGRL-LAMEDDLPYVOIT 273
DB 1032 ITGLPRDPSDIPFNSSTGELVDPSTGKPIINNSTAGIV--SGKPGGLPIEDE----- 1081
QY 274 PNGDLKTVGRFD-----FDQLESTYMAHPRVDESGELFALSYDVVSKPYLKYFRFSP 327
DB 1082 -NGNL-----FDPSTNLPIDGN-----NQLVNPETNSTGSGTSGTKRP-----KP 1121
QY 328 -----DGTSPDYET-QLDQ-----PMMHDFAITENFVVPVDDQVVF 364
DB 1122 GIYPVNGGVVPPDEAKDQADKGDGLIYPPTNSINKPDTNTQYNSNTGNTIINPETGKY- 1180
QY 365 KLEPMIRG-----GSPVYVDKNKVARFGILDKYAEDESNIKMIDAPCCFCHLMNAME 417
DB 1181 -IPGSLPGSLNPSFNFPQOTDE-----ITGKPYDVTYGLPY----- 1216
QY 418 EPEETDEVVYIGSCMPDPSIFNESDENLKSYLEIRLNKLTGSESTRPPIISNEDQVNL 477
DB 1217 DSTGTGLIDPATKRLPIPGSVAD-----EILLEV-LNTITDEVYGLPI-----DLE 1261
QY 478 AGVNNRNLGKRTKFAYLALAEPMKVSFAKY--DLTTEGVYKHLKYDNRNGEPLFL 534

DB 1262 TGL-----PRDPVSGLPOLPNCGLTDPNSKKRIPGS--HSG---FI 1297
QY 535 PGEGEE 541
DB 1298 NRTSGEQ 1304

RESULT 6
US-08-928-361B-11
Sequence 11, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vermy, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1042 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-11

Query Match 3.7%; Score 116.5; DB 3; Length 1042;
Best Local Similarity 19.9%; Pred. No. 0.0098;
Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;

QY 48 LNVSSALHTPPPLHFPKSSNSPAL---VVKPKAKESNTKQNL----- 88
DB 190 LNPATGVMLPGSLGPSEQTFPSPEIDGIIIPPEVAANADKFKLSIPSPVESIPEKDK 249
QY 89 -FORAAMALDAEGLVSHKHLPLPKTADPSVOIAGNFAPVNEQVRRNLPPVYGLPD 147
DB 250 KIDSISELMYDIESGRILQGVSKRIPGS-----IAGDLNPKIPTQTD-SYTGKRPID 302
QY 148 SIRGVYVRNGANPLHEPVYGHFFPDGDGVAHAKFEHGSASY--ACRFOTNRFOEROL 205
DB 303 PTTGL-----PFNRP-TGH-----LINPTNNNTMDSFAGAYKVAVNSGIKTDNVY 347
QY 206 GRPV-----FPAIGELGHGTIARLMLFYAARAAGIYDPAHGTGVAN-- 248


```

1      RESULT 7
2      US-08-928-361B-5
3      ; Sequence 5, Application US/08928361B
4      ; Patent No. 6071518
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Petersen, Carolyn
7      ; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
8      ; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
9      ; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
10     ; TITLE OF INVENTION: SPECIES INFECTIONS
11     ; NUMBER OF SEQUENCES: 30
12     ; CORRESPONDENCE ADDRESS:
13     ; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
14     ; STREET: 385 Sherman Avenue, Suite 6
15     ; CITY: Palo Alto
16     ; STATE: CA
17     ; COUNTRY: USA
18     ; ZIP: 94306-1840
19     ; COMPUTER READABLE FORM:
20     ; MEDIUM TYPE: Floppy disk
21     ; COMPUTER: IBM PC compatible
22     ; OPERATING SYSTEM: PC-DOS/MS-DOS
23     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
24     ; CURRENT APPLICATION DATA:
25     ; APPLICATION NUMBER: US/08/928,361B
26     ; FILING DATE: 12-SEP-1997
27     ; CLASSIFICATION:
28     ; PRIOR APPLICATION DATA:
29     ; APPLICATION NUMBER: US 60/026,062
30     ; FILING DATE: 13-SEP-1996
31     ; ATTORNEY/AGENT INFORMATION:
32     ; NAME: VERNY, Hana
33     ; REGISTRATION NUMBER: 30,518
34     ; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
35     ; TELECOMMUNICATION INFORMATION:
36     ; TELEPHONE: 650-324-1677
37     ; TELEFAX: 650-324-1678
38     ; INFORMATION FOR SEQ ID NO: 5:
39     ; SEQUENCE CHARACTERISTICS:
40     ; LENGTH: 1837 amino acids
41     ; TYPE: amino acid
42     ; STRANDEDNESS:
43     ; TOPOLOGY: linear
44     ; MOLECULE TYPE: protein
45     ; US-08-928-361B-5

```

QY	48	LNVSALHTPPLALHPKROSSNSPAI----	VWPRKAESTXKMNL-----	88
Db	985	LNPRATGMATGSLGSPSQTPTFSEIEDGCI	TPPEVAANAADKFKSLTSPSPSESTPEKDQ	10444
QY	89	-FORAAALDAAEGLVSHKHLPLPKTADPSV	IOAGNFAPVNBQPYRRLPVVGKLPD	147
Db	1045	KIDISISLAWDIESGRILIGVSKRPIPGS----	IAQLANITMKTPQTD-SVYGRKD	1097
QY	148	SIGGVYVRNGANLPHBPVTGHHFFDDG	NAVHAKKEHGSAST--ACRFTQTRFVQEQOL	205
Db	1098	PTTGL-----PFPNP-TGH-----	LINPTNNNTMDSFAGAYKVAVSNKIKTDVY	1142
QY	206	GRPV-----	FPKAIHELHGHHTJARLMLTYAANAAGIYVDP	PHAGTVAN--248
Db	1143	GLPDEITGLKDPVSDIPNSTTGL-----	-----VDPSTGKAPINNT	1182
QY	249	AGLVFNGRLIAESDDLPRQVOTITENGDL--	KTVGRFPDGOLESTWIAHPKVPESG	305
Db	1183	AGIY-SGRKRGPIEDE-----	NGNLPDSTKLPIDGNNOJ-----	VNPETN1223
QY	306	ELFALSYDVNSKPYLKIYFRSP-----	DGTSPDVEI--OLDQ-----	P342
Db	1224	STVSGSTSGSTKTP-----	KPGIPVNGGCVAPDEEAKQOAKGKDGLIVP	PTNSINKDP1276
QY	343	TMMHDEPATEFVVPVDPQVYFKLPWEMIRG	-----GSPVYVDKNKVARFELDKYAD	395
Db	1277	VTNTQYISNTGNINIPETGKV--IPGSLP	BSLNPSEFNPOQTD-----ITGKPVDT	1327
QY	396	SSNTKWIADAPDCFCFHLMAABEPENDE	EVVVISQCTPPDSIFNSEDENKLSVLSERLN	455
Db	1328	VTGLPY-----	DPSTGEIITDAITKLPIGVSAGD-----	EILFV-LN1364
QY	456	LKTGESTRPPIISNEDQVNLKGAVMYNRML	GKRTKFAYLALAEWPVKVSGFVAV--DL	512
Db	1365	ITTDEVATGLPI-----	DLFETGL-----PRDPVSGLPQLPNCGL	1397
QY	513	TTGEVKKHLYGDMRYGGEPLFLPGEEGEE	541	
Db	1398	VDPSNKKRPIPS--HSG--	FINGTSGEQ1421	

```

RESULT      8
US-09-461-697-77
; Sequence 77, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: LO, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Purnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 77
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-77

```

Query Match 3.4%; Score 106; DB 4; Length 412;
Best Local Similarity 20.6%; Pred. No. 0.024;
Matches 74; Conservative 45; Mismatches 131; Indels 110; Gaps 16;

QY 23 PLSSSSSSSLSTCSSLPMASTRKLNSSALHTPPALHPKQSSNPAIVYKPKAK--- 79
DB 9 PLSSSOTNHSF-DLLPREFRLVE-----VHDP-LHQPANKPKPTMDIDISEPCS 59
QY 80 -----ESTWKQWNLFORAAALDAAEGLVSHKTLPLPKTADSVQIAGFAFV 130
DB 60 LTIHTIQLQHRRLRLATQAQNOOOTE--VKTESEPL-SCPSPLPLDLDLPL 116
QY 131 N-EQVRRNLPLYVGLKPLDSIKGVYVNGANPLHEPV-----GHHF 170
DB 117 DCKNP---NAPFOIRHSDESDPYRGKG-----EPVTELSMWSCHROLLYQAVATILAHAG 168
QY 171 FGGDGVAVHVKFEHGSASACRFQTQNRVQERO--LGRPVYKPAIGELHGTGTLARML 228
DB 169 FDCANESVLETLDVAHEXCLKFTKLRLFAVDREARLGGTPPDVDEOVFEHVGISVLS 228
QY 229 F-----YARAAGIVDPAGTGVANAGLVYFNGRLIAME 263
DB 222 LCKFMQHRIKDYHMYLQISKQLESEYER---INPEKAT-----E 266
QY 264 DDLPTQVQITPNGDLKTYGRFDPDQLESTMTAHPKVD-----ESGELFALSYDVVSKP 318
DB 267 DAKPVAKIKEEPVSDI---TFPVSELEADLASGDQLPMGVLAGAOSERFPNLEVEASP 322

RESULT 9
US-08-611-107-31
; Sequence 31, Application US/08611107
; Patent No. 5801233
; GENERAL INFORMATION:
; APPLICANT: Haselkorn, Robert
; TITLE OF INVENTION: NUCLEIC ACID COMPOSITIONS ENCODING
; TITLE OF INVENTION: ACETYL-COA CARBOXYLASE AND USES
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611.107
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US SN 07/956,700
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/09340
; FILING DATE: 30-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US SN 08/422,560
; FILING DATE: 14-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: ARCD:221
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:
; LENGTH: 2172 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-611-107-31

Query Match 3.3%; Score 104.5; DB 1; Length 2172;
Best Local Similarity 19.9%; Pred. No. 0.65; Mismatches 247; Indels 147; Gaps 27;
Matches 118; Conservative 80;

QY 66 SSNSPAIVYKPKAKESNTKQNLFORAAALDAAEGLVSHKTLPLPKTADSVQIAG 125
DB 1406 TGNCTQVDIYREVEDPNTHK--LFYRSATPPAGPLHG--IALHEPKPL-----DAIDLKR 1457
QY 126 NFAPVNEQDVNRNLVVGKL-----PDSIKGVYVNGANPLHEPVYTHGHHFDDGDM-- 176
DB 1458 AARKNETTYCYDFPLAETALKKSMESGISVABSNHNOFYAEVTELIFADSTGSGWT 1517
QY 177 -VHAKFEHGSASVACRFQTQNRVQERQLGRPV-----PRAIGELHGT 221
DB 1518 PLVPERPPGSNNGVYVANNKLSLTPPEPGRELLIVANDVTFKAGSGPRDPAFDPAVT 1577
QY 222 GIA--RLMLFYARAAGIVDPAGTGVANAGLVYFNGRLIAMESDLP---YQVQITP 274
DB 1578 NLACRKRIPLIYLSATVGA---RLGVAEEIKACFH--VGMSDQSPERGPHYIYLTE 1629
QY 275 NGDLKTVGRFPDQLESTMTAHPKVDPESGELFALSTDVYK----- 317
DB 1630 Q-----DY-SRLSSSVIAHELKVPESGTRFWVVDIYVKGEDGLCCENLHSGAIAS 1679
QY 318 PLYKYFR--FSPDGTKSPDVEI-----QLDPTMMDHFAITENPVVYVDDQV 362
DB 1660 AYSKAYRETFTIETVGAIGATLARLGARCIORLDQPIILIGYSALKNL---GREV 1736
QY 363 VKLEPMIRGGSPVYVYDKNKAARFGILDKVAEDSSNIMIDAPDCCFPHMANNEPEPTD 422
DB 1737 YSSQQL--GGPKIMATNGVYHLVSDLDGVSAILKWL-----SYVPYVG 1781
QY 423 EYVYVIGSCMTPPDSIFNSDENLKSVLSEIRLNLTGSGTRPIISNDOQVNLGAVVN 482
DB 1782 GPLPTVKSLDPERAVTYPEP-----SCDARAALICGIDDTGKMLSGMFD 1827
QY 483 RNMLGRKTKFAVLALEPPPK--VSGFAKVDLTGGEVKKHLYGDNRYGGEPLFLPG---- 536
DB 1828 RE-----SFEVLEBGMATVITIGRAKSAKTAQA---LLDFNR-BELPLFIAMNRG 1875
QY 537 -EGGED--EGY---LCFVHDEKTKSE--LQIVNAVSLVEAVTVKLPSRV 580
DB 1876 FSGGQRDLEFGILQGXMIVENLRTYKOPAFVYIPKAGELRGAVVVDVSKI 1927

RESULT 10
US-09-001-984C-106
; Sequence 106, Application US/09001984C
; Patent No. 6245331
; GENERAL INFORMATION:
; APPLICANT: Laal, Suman
; APPLICANT: Zolla-Pazner, Susan
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
; FILE REFERENCE: NYU-011
; CURRENT APPLICATION NUMBER: US/09/001.984C
; PRIOR FILING DATE: 1997-12-31
; PRIOR APPLICATION NUMBER: 60/034,003
; PRIOR FILING DATE: 1996-12-31
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 106
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis

US-09-001-984C-106

Query Match 3.1%; Score 98; DB 4; Length 741;
Best Local Similarity 21.7%; Pred. No. 0.47; Indels 208; Gaps 37;
Matches 139; Conservative 79; Mismatches 216; Indels 208; Gaps 37;

```
QY 97 LDAAGFLVSEHKLPD-----KTADPSVOJAGNFAVNEQPV---RRNLFPV---GK 144
DB 78 MDAVRQFLTEIGYLLPEPDDEFTITSGVDAEITTAGQOLVYPVLANAFALNAARMS 137
QY 145 LPDSIKGYV---RNGA---NPLHEPVTC-----HHPFDGGMVHAKFEHGSAYACR 192
DB 138 LYDALYGTVDIPEFDGAEKGPYTNKVRGDKVLAARKEFLD-----DSVPLSGSGFGDAG 192
QY 193 FTQNRFRVQEOHOLGRPVPRKAI-----GELHGTGIAR-----LMLFARAAGIVDPAH 242
DB 193 FT-----VQDQOLVVALPDKSTGLANPGQFAGYGAESPYSVLLINGLHLEILLIDPES 247
QY 243 GTGVAN-AG-----LVYFNGRLAMSEDD--LPYOVQITPN-GDLKTV----- 281
DB 248 QVGTIDRAGVNDVILESAITTIMDEDSVAANDADKVLGYTNMGLKNGDLAAAVDKG 307
QY 282 -----GRDFEDGQ-----LESTMIAPKVPDESGELFALSVDYVS 316
DB 308 TAFLEVLNRDRNYTAPGGGQFTLPGRSLMFVRNVGHLMTNDATVDTPDSEVEGIMDALF 367
QY 317 KPYLKYFRFSDGTRKSPV-----ETOLDPTMMH---DFAIT-ENFVVVPPDOV 362
DB 368 TGLI-----AIGHLKASVNGPLINSRTGSITYVR-KHGGAEVAFCELSRVED--- 418
QY 363 VFKEIPEMIRGSPVYYDKNKVARFGILDKYAEDESSNIKWIDAPDCECFHLMNAMEPETD 422
DB 419 VLGLPQ-----NTMKIGIMBERTYVNLK-----ACIKKAAD 451
QY 423 EVVYVIGCMTPPDSITFESDENLKSVLSEIRLNLTGSTRP-----IISNEDQOVN 475
DB 452 RVFEINTGF-----LDRTGDEIHITSMAGRPVGRGTMSQPMILAYEDHND 498
QY 476 ---LEAGMVRNMLGR-----KTKFRAY--LALAEW-FKVSQ-----FAKY 510
DB 499 AGLAGFGSRGQVQKGMWMTMELMADWETKIAQPRAGASTAMVSPPTAATLHALHYHOV 558
QY 511 DLTGEVVKHLYGDNRYGGEPL-----FLPEGGEDEDE-----GYILCFVHDE 553
DB 559 DVAA--VOGGLAGKRRAITBOLLITPLAKELAMAPDEIREVDNNOCSILGIVVMV-DQ 615
QY 554 KTWKSEIQIVNAVSL-EVEATVKLPSRVYGF-HGFTIGAD 592
DB 616 GVGCSKVPDIHDVALMEDRATLRISSQLLAWMLRHGVITSAD 657
```

RESULT 11

US-08-097-829-4

Sequence 4, Application US/08097829

Patent No. 5498831

GENERAL INFORMATION:

APPLICANT: Burgess, Diane G.

APPLICANT: Dooner, Hugo K.

TITLE OF INVENTION: Pea ADP-glucose Pyrophosphorylase Subunit

TITLE OF INVENTION: Genes and Their Uses

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Townsend and Townsend Kourie and Crew

STREET: One Market Plaza, Stewart Tower, Suite 2000

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

```
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/097,829
FILING DATE: 19930723
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 12176-35
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-097-829-4
```

Query Match 3.1%; Score 97; DB 1; Length 516;
Best Local Similarity 20.9%; Pred. No. 0.32;
Matches 123; Conservative 69; Mismatches 172; Indels 224; Gaps 32;

```
QY 1 MASFTATAVAGRWLGNHTOPLSSOSSDLSYCSLPMARVTRKLVSSA-----L 54
DB 1 MASMAIGYVK-----VPPSSSSSS-----SSSSSAIAIRNLSTFSQCGDKI 46
QY 55 HTTPALHPFKOSSNBPAYVVKPKA-KESNTKQMLFORAAAALDAAGFLV---SHEKL 110
DB 47 FT---VSGTRSSGSRNRPFLVSPKAVSDSKNSQTCIDPDASRSVL---GILGGAGTRL 99
QY 111 HPL-FKTADPSVQIAGNAPVNEQVRRRLPVYVKIPDSIKGYVRNANGNPLHEPTGHH 169
DB 100 YPLTKRRAKPAVPLGANYRLI-----DIPVSNCLINSMSIKIYVLTQFN---SASLNKH 149
QY 170 F-----PDGGMVAVAKFE-----HGSA---SYACFQTQTN----- 197
DB 150 LSRATAVSNIGYKNGSFVYVLAQOSPENPNPFGTADAVRQYLMFEEHNVLEYIVLAG 209
QY 198 ---RFVDERQUGRPVFPKAIQELGHGTGIARLMLFYARAAAGIYDPAHGTGVANAGLY 253
DB 210 DHLVHMDYER-----FLQAHRESADITVASLPMDEARATA-----FGLMK 250
QY 254 FN--GRLAMSE-----DD-----LPYOVQITPN-----D 277
DB 251 IDEEGRIYFSEKPRGEOLKAMKVDTTILGLDERAKEMPY---IASMGIYVVSKHVMD 307
QY 278 L---KTVGRFEDDGQLESTMIAPKVPDPESGEL---FALSYD----- 313
DB 308 LIRDKFPGANDFGSE-----VTPGATELGLRVQVLYLDGWEDIGITEAFNANL 357
QY 314 -VVSKP---YLKYFRFSDGTRKSPVVEIQLDQPTMWHDFAITENFV---VVPDOGV--- 362
DB 358 GITKKVPDPFSYDSSPIYTOP---RYLPPSKMLDADITDSVIGECVYKKNCKIHS 412
QY 363 VKLPEMT-----RGSSPVYYDKNKKARIGILDKYAE 394
DB 413 VVGLRSCISEGAIIEEDTILMGADYETDARRFLAKGAGVPVIGIGNSHIKRAIIDKNAR 472
QY 395 DSSNIKWIDAPDCECFHLMNAMEEPE---TDEVVVISCMTPPDSI 437
DB 473 IDDDVKIINSD-----NVOEAAREHETGKIGIYIVINDALIPSGTV 515
```

RESULT 12

US-08-577-403-4

Sequence 4, Application US/08577403

Patent No. 5773693

GENERAL INFORMATION:

APPLICANT: Burgess, Diane G.

APPLICANT: Dooner, Hugo K.

TITLE OF INVENTION: Pea ADP-glucose pyrophosphorylase Subunit
TITLE OF INVENTION: Genes and Their Uses
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend Kourie and Crew
STREET: One Market Plaza, Steuart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/577,403
FILING DATE: 22-DEC-1995
CLASSIFICATION: 800
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: 08/097,829
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 12176-35
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-577-403-4

Query Match 3.1%; Score 97; DB 1; Length 516;
Best Local Similarity 20.9%; Pred. No. 0.32;
Matches 123; Conservative 69; Mismatches 172; Indels 224; Gaps 32;

QY 1 MASFTATAAVSGRWLGNGHTOPPLSSSGSSDLSYCSLPMASHVTYRKLNVSSA-----L 54
DB 1 MASMAAIGYK-----VPPSSSSSSSS-----SSSSSKAIARNLSFTSQCGDKI 46
QY 55 HRPALHPFKOSSNSPAIYVVKPKA-KESNTKQKNLFQRAAAALDAABEFLV---SHEKL 110
DB 47 FV--VSGTRRSSGRNPFLVSPRAVSDSKNSQTCIDPDASRYL--GILGGGAGTR 99
QY 111 HPL-PTADPSVQIAGNFAPVNEQPYRRMLPVYVKLPDSIKGVYVNGANPLHEPYTGH 169
DB 100 YPLTKRKARPAVPLGANYLI-----DIPVSNCLNSISIKIYVLTQTN---SASLNH 149
QY 170 F-----FDGDGMVAHVAKFE-----HGSA---SYACFTQTN----- 197
DB 150 LSRAYASNLGKYKNEGFVLAQQOSPENPNWFOGTADAVROQLMFEHNVLLEYVL 209
QY 198 ----RVOEQRLGRVFPRAIGELHGHGTIARLMFLYAAAGIYVPAIGTGAANGLY 253
DB 210 DHLRYRDYR-----FIQAHRESADITVASLPMDEARATA-----FGLMK 250
QY 254 FN--GRLLAMSE-----DD-----LPYQVQTPNG-----D 277
DB 251 IBEGRIVFSEKPKGEQJLAKMKVDTTILGLDDERAKEMPY---IASMGIIYVYSKHVMD 307
QY 278 L--KTGVGFEDGQLESTMIAPKVDPESGEL---FALSVD----- 313
DB 308 LRDKRFPGANDGSE-----VIRGATELGRLVQYVLVDGVMEDIGTIEAFYANML 357
QY 314 -VYSKP---YLKIFRSFPGTSPDVEIQLDQPTMAHDAITENYV---VPPDOYV--- 362
DB 358 GITKKRPVDPFSFYDRSSPIYTOP---RYLPPSKMLDADITDSVIGECVINKCKTIHNS 412

QY 363 VEKLPEMI-----RGSSPVYVDKNKVARREGILDKYAE 394
DB 413 VVGLRSCISEGAIIEDTLTMGADYETDADRFRFLAAKGVPGLIGKNSIKRAIIDKNAR 472
QY 395 DSSNIRKWDAPDPCFHLNMAWEEPE----TDEVVIGSCMTPPDSI 437
DB 473 IGDDVXKINSND-----NVQEAARETEGYFKSGIYVIVIKDALIPSGTV 515

RESULT 13
US-09-335-409-3
; Sequence 3, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1410
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-3

Query Match 3.0%; Score 96; DB 3; Length 1410;
Best Local Similarity 23.2%; Pred. No. 2.4;
Matches 107; Conservative 57; Mismatches 182; Indels 116; Gaps 25;

QY 10 VSGRWLGNGHTOPPLSSSGS-SDSLST-----CSSLPMASRYTRK-----LNVSSALHT 56
DB 645 VSEAGYEGSGDQPPMPPIOTPSDLAVITYTSSGTGLPKVMTDHRKAVVTIIDINERFEI 704
QY 57 PALHPFKOSSNS-----PAIYVVKPKA-KESNTKQKNLFQRAAA---AL 97
DB 705 GGEDRVLAISLSFELSYDVFEGILAAAGTIYVPOASKLRDPAHMAELLERKRVYWN 764
QY 98 DAEGFLVSHKLIHP--LPKTADPSVQIAGNFAPV---NEQVYRNLPVY---GKLPS 148
DB 765 PALMRVLVEHFGRPDLSARLSRLSL--LSGDMIPVGLPGLPELOAIRGVSVISLGATTEAS 823
QY 149 I--KGVYVNGANPLHEPYTGHFFDGDGMVAHVAKFEHGSASVACRF--TQTRRFVQERQL 205
DB 824 IMSIGIPVKN-----VDSLMASTPYGRRLRQGTHTVLDLEALE 860
QY 206 GRPVPFKAIGELH-GHTGIARLMFLYAR-----AAAGIYVPAHGTGVANAGLVYENGRL 259
DB 861 PRPVW--VPGQLYIGGVGLA---LGWRODEKTRKSFVLVHPETGERLXYTGDL---GRYL 912
QY 260 AMSEDDLPIYQVQITPBGDLKTYGRFPD-----GQLSTIMAHKXVDPESGELFA 309
DB 913 -----PDGNTIEFPGREDNQIKLGRYRVELGEIEETLKSHPNV--RDAVIYR 956
QY 310 LSYDVYKPYLYKFRSPGTSKSPDVEIQLDQPTMAHDA---ITENFVYVPPDOYVFKLP 367
DB 957 VGNDAANKLLIAY--VVEGTRRRRAEODASIKTERITDRRAHAAEDGLSDGERVOFKLA 1014
QY 368 EM-IR---GSPVYVDKNKVARFGILDKYAEDSSNIKWIDAP 405
DB 1015 RHGIRRDLDGKRPVVDLTGQDPREAGLDVYARRRSVATFLEAP 1056

RESULT 14
US-09-568-102-3

